Frog Counting Tool

NIEK ZIEVERINK, University of Twente, The Netherlands

Fig. 1. Two frogs sitting on a leave

Frogs are a good indicator species for the health of an ecosystem. For this reason, the size of the frog population in a certain ecosystem can give a good indication of the health of that ecosystem. At the same time, it is harder to estimate the size of a frog population than it is to estimate the size of other populations of animals. Therefore frog populations are often not represented well in existing Species Distribution Models (SDMs). In this paper, a model is proposed that is able to more accurately calculate the distributions of frogs in an area. In the first part of this research, a review is given of the existing literature in the field. In the second part, an actual model is proposed to effectively predict frog distributions.

Additional Key Words and Phrases: Frogs, indicator species, ecosystem health, Species Distribution Model

1 INTRODUCTION

Biodiversity and ecosystem health are obviously very important to our survival and to the world in general. Healthy ecosystems provide us with many services. [4] Divides these services into four categories, each of which is crucial for our existence or well-being:

(1) Provisioning services. Ecosystems provide us with all kinds of things like food or fuel.
(2) Regulating services. Ecosystems regulate the environment, for example by purifying our air and water.
(3) Cultural services. Ecosystems provide us with services that can improve our well-being (think of forest walks for example).
(4) Supporting services. Ecosystems also provide supporting services, which make all other services possible.

Unfortunately, ecosystem health has come under more and more under pressure lately, mostly because of human actions. According to [4], we are now facing the greatest species extinction rate since

---

TScIT 37, July 8, 2022, Enschede, The Netherlands

the event that wiped out the dinosaurs (and many other species) 65 million years ago.

It goes without saying that the destruction of ecosystems would have catastrophic consequences. It is therefore important to actively reduce ecosystem destruction and even improve ecosystem health if possible. The first step towards improving ecosystem health is to do regular measurements. By doing so, researchers can learn about the current state of an ecosystem and see whether their actions are having the wanted effect.

The problem is that 'ecosystem health' is an abstract concept that is not easy to measure. To get an indication of the health of an ecosystem, researchers therefore often use so-called indicator species. Indicator species are organisms that can serve as a measure of the environmental conditions in a certain area by their presence in that area [2].

Researchers can then build a so-called Species Distribution Model (SDM) to estimate the population size of such a species, giving them an indication of the ecosystem health. [6] defines an SDM as a "model that relates species distribution data (occurrence or abundance at known locations) with information on the environmental and/or spatial characteristics of those locations . . . The model can be used to provide understanding and/or to predict the species' distribution across a landscape." (p. 678).

So to train an SDM, information about the distribution of a species in certain locations is combined with other environmental data in those locations (e.g. the temperature). From that other environmental data, variables can be derived which can be used to predict the population size of the species in other locations.

According to [10], frogs are good indicator species because their skins are very permeable, making them highly susceptible to pollutants. This means that if ecosystems get polluted, frogs will disappear. The size of the frog population in a certain area is therefore a good indicator of the pollution of that area. At the same time, frogs are often poorly served in existing SDMs, because they tend to live in a very localized area [10].

For these reasons it is relevant to build a new model which can more accurately predict the density of a frog population in a certain area. This is the goal of the EY Data Challenge of 2022 [9] and also the goal of this research.

2 RESEARCH QUESTIONS

This research is conducted in order to complete level 3 of the EY Data Challenge of 2022 [9]. The objective is to build a Species Distributino Model to predict the density of frog population for regions covering Australia, South Africa and Costa Rica. Before the model can be used to accurately estimate frog population density, a number of predictor variables should be derived from the datasets that the model is trained on. Therefore the following two research questions will need to be answered:

(1) What predictor variables can be used to accurately indicate the size of a frog population?
(2) Can a model be built that can accurately predict the size of a frog population using certain predictor variables?

3 LITERATURE REVIEW

The first part of this research consisted of a review of the existing literature on using Species Distribution Models for frog populations.

To start off, [6] gives a good introduction into the field. As was already mentioned in the introduction, Elith defines an SDM as a "model that relates species distribution data (occurrence or abundance at known locations) with information on the environmental and/or spatial characteristics of those locations . . . The model can be used to provide understanding and/or to predict the species' distribution across a landscape." (p. 678).

3.1 SDMs for frogs

Let's continue with the existing literature on SDMs for frogs. This research is relevant because it gives insight into which kinds of data have proved to be working well for predicting frog population sizes.

[15] combined an SDM with species co-occurance patterns to create a Joint Species Distribution Model for predicting frog density. The benefit of this is that this model can also show whether the distribution of two different species are related to eachother.

3.2 SDMs for amphibians

Because research on SDMs for frogs in particular has been quite sparse, research on other kinds of amphibians was examined as well. The assumption was made that variables that would accurately predict the distribution of these species would help do so for frogs as well because amphibians and frogs are quite similar.

[12] Used satellite data along with data from GIS systems to predict the presence of amphibian species (under which were frogs) near water reservoirs. This data was used for training an Artificial Neural Network model, which managed to achieve an accuracy of 100% for its task. This research will also use satellite and GIS data.

[11] Aimed to find out how vulnerable populations of amphibians (under which were frogs) would be to climate change and the higher temperatures as a consequence of this. Historical climate data was used to train an SDM that could predict the future distribution of the amphibians, given the expected increase in temperature. Temperature would be a good predictor variables for this research as well.

[16] Showed the importance of comparing multiple models for predicting the frog presence. This is something that might be interesting to do for this research as well. Even though the original idea for the EY Data Challenge is to use a linear regression model, it would be interesting to compare this to other models to see which one would be the most effective.

3.3 Transferability

Lastly, after the predictor variables have been decided on, the bias in the data is addressed and the model is built, it should be able to predict frog population sizes in Australia, Costa Rica and South Africa. Therefore it is a good idea to also consider research on the transferability of SDMs.

[17] Shows that different predictor variables lead to more or less transferability of the model. Some variables are more transferable than other variables. This shows that it is important to test each

of the predictor variables that will be used in the model for its transferability across the other regions.

[5] considers transferability across Spain and Australia. This paper is particularly interesting, because Australia is one of the regions used in this research as well. In the paper, multiple algorithms are compared and the conclusion is drawn that General Linear Models and Generalized Additive Models have the best performance and transferability. It would be interesting to test these algorithms for the regions covered and compare their performance to the simple linear regression model used initially.

3.4 Lessons learned

The literature review has shed some light on what SDMs are in general, how they can be modeled and how they can be made transferable across different areas. It has also uncovered some predictor variables, like the temperature and GIS-data, which have proven to be effective. This has partly answered the first research question. The next step will be to build the actual model, which will likely uncover some additional predictor variables and will also help answer research question two.

4 BACKGROUND INFORMATION

Before the methodology used for building the model is explained, it is important to have some background knowledge. In the following sections, first the modeling process of SDMs is explained. Then, the concept of pseudo-absences is introduced and it is explained how this could help mitigate bias in the data.

4.1 Modeling

Figure 2 [14], gives a good overview of how the Species Distribution Modeling process works in practice.

The first two steps are to gather observation data (for the target variable) and environmental data (for the predictor variables). Then, the datasets can be processed and combined, after they can be split into a train and a test set used for training and testing the model.

The next step consists of choosing the model algorithm and actually training it on the train data. The resulting model can then be used to predict the frog distribution in other areas.

4.2 Pseudo-absences

Usually, there is only presence data for species. This makes sense, because it is easy to log an observation of a certain species, but it is much harder to tell whether a species is absent in a certain area. According to [14], this lack of absence data is a big source of bias in SDM models.

If there is only frog presence data available, each area where no observation has been recorded could be seen as a potential absence for that species. But the problem is that it is not clear whether the species are truly absent there, or if it just has not been observed.

This bias can be addressed by generating 'pseudo-absences'. Pseudo-absences are generated from all the potential absence points and can be used to replace actual absence data. According to [3], pseudo-absences do indeed improve model accuracy if proper absence-data is missing.

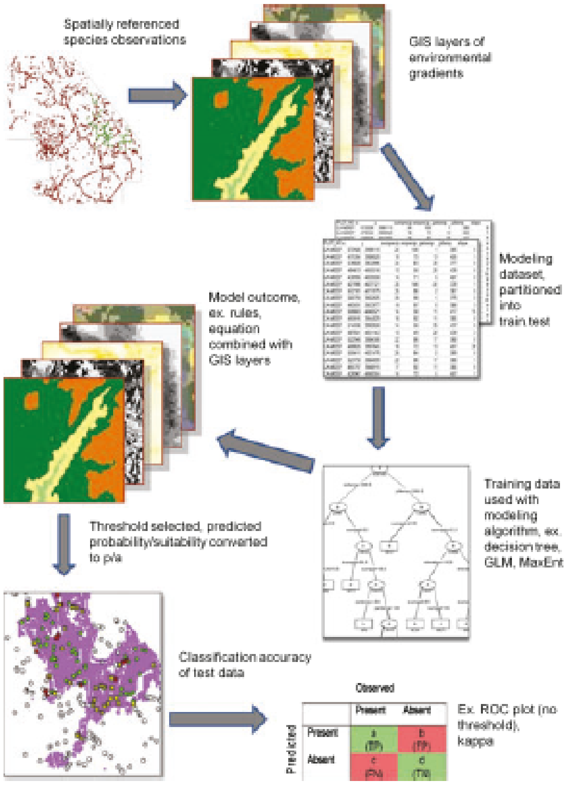

Fig. 2. Miller

There are multiple ways to generate pseudo-absences. For this research, the method that was selected is the one provided by [19], which is a three-step method based on a couple of assumptions.

Firstly, it assumes that the probability that a species is present in an area but has just not been spotted is higher if that area is environmentally similar to the area of a known occurrence.

Secondly, it assumes that this probability is also higher if the area where no species have been spotted is geographically close to the area of a known occurrence.

Based on those assumptions, pseudo-absences are generated in three steps. First of all, a set of potential absences data is generated based on all the locations where no presence point has been recorded.

Secondly, the data points that are environmentally similar to the known occurrences are removed from this set. This can be done as follows. First, a clustering algorithm is trained using the presence data. Then, the absence dataset is fed to that model. All absence data points that are too environmentally similar (i.e. too close to the center of a cluster) to the known presence points are then removed from the set.

Lastly, the coordinates of all the data points in the remaining dataset are compared to the known presence points. All the values that are geographically too close to the known presence points (i.e. their coordinates are too close to the coordinates of any known presence points) are also removed from the set. The remaining set

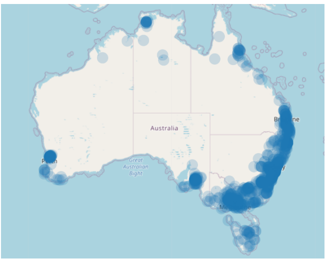

Fig. 3. Frog presence data - Australia

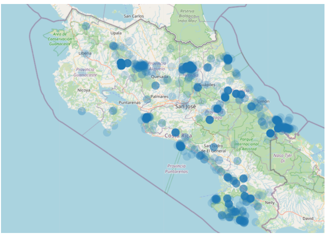

Fig. 4. Frog presence data - Costa Rica

of pseudo-absences are considered likely to be true absences and are therefore added to the dataset that the SDM will be trained on.

5 METHOD

For the data challenge, EY provided two datasets containing frog presence data. There was one dataset for the region of Australia and one dataset for the regions of Costa Rica and South Africa. These datasets were combined and the latitude and longitude of each frog presence was extracted.

EY also provided a number of base models. These models were used as a starting point and were then improved upon to make the models perform more accurately. The method for doing so is explained in the next sections.

5.1 Area selection

As can be seen in figures 3 - 5, large parts of the countries in the provided dataset contain barely any frog presence data, if at all. The parts that do not contain any data are not relevant for training the model and only take up valuable computing resources.

For this reason, sub-areas of data were selected within these countries for the data generation. This was done manually, by looking at the frog presence data in the countries (see figures 3-5) and selecting areas that have a high density of frog occurrence data.

5.2 Grid creation

After loading the data, the frog data was cut into equal sized grids of 5 degrees longitude and 5 degrees latitude (about 5 square kilometres [1]).

Each grid got assigned a frog count, based on the number of frog occurrences within the area of the grid. The frog count was used as the target variable for the model.

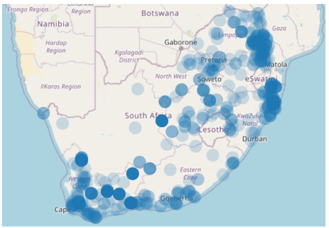

Fig. 5. Frog presence data - South Africa

Table 1. Datasets used

dataset	type of data
TerraClimate	climatic data (e.g. minimum temperature, maximum temperature)
Sentinel-2 Level-2A	satellite data (i.e. color bands; e.g. red, green, blue)
Copernicus DEM GLO-30	elevation data
JRC Global Surface Water	surface water data (e.g. frequency with which water returns yearly)
Esri 10-Meter Land Cover	land cover data (per pixel; e.g. 'forest' or 'water')

5.3 Predictor variables

A number of datasets, all provided by the Microsoft Planetary Portal [13], were used for extracting the predictor variables. Table 1 provides a summary of the datasets that were used.

Each of these datasets contains multiple features. The value of each of the features for a grid was calculated by averaging all the values for that feature within the area of the grid. The resulting values were then added to the dataset to be used as predictor variables in the model.

5.4 Pre-processing

Before training the model, a number of pre-processing steps were carried out to increase model performance.

5.4.1 Data cleaning. Firstly, the data was cleaned up. NaN values were removed from the set, as well as outliers (grids with exceptionally high values for the frog count). The outlier removal was done manually by plotting all of the frog counts and choosing a limit above which datapoints would be removed. This value depended on the data fed to the model, but was usually around 100-200.

5.4.2 Mapping categorical values. The datasets contained some categorical values, which used an integer to represent a certain category (e.g. the land cover value provided by the Esri dataset). These values were converted to binary format using one hot encoding.

5.4.3 Normalizing data. Some predictor variables have larger variables than others. To mitigate the impact that these higher values have on the model performance, a simple min-max normalization

Table 2. Predictor variables

Variable
Maximum temperature (tmax)
Minimum temperature (tmin)
Accumulated precipation (ppt)
Soil moisture (soil)
Normalized Difference Vegetation Index (NDVI)
Elevation
Slope (in %)
Land cover
Frequency with which water was present (occurrence)
Frequency with which water returns from year to year (recurrence)
Number of months that water was present (seasonality)

algorithm was applied to make sure each value would be between 0 and 1, depending on their relative size.

5.4.4 Feature engineering. From the gathered features, a number of new features were engineered. These features were thought to be better at predicting the species distribution than the raw features. The features that were calculated were:

- Normalized Difference Vegetation Index (NDVI): (NIR - Red) / (NIR + Red) where NIR is the value of the NIR (near infrared) band provided by the Sentinel-2 dataset and RED the value of the red band. This gives an indication of the vegetation health in a certain area [7].
- Normalized Difference Water Index (NDWI): NDWI = (Green – NIR)/(Green + NIR) where NIR is again the value of the NIR band and Green the value of the Green band, both provided by the Sentinel-2 dataset. This gives an indication of whether there is open water in an area [8].
- Slope. The slope of the elevation in an area (in %). This was calculated using the elevation data provided by the Copernicus dataset.

5.4.5 Feature selection. To prevent unnecessary complexity, only the most effective features were eventually included in the model. The features that were deemed to be most important for model performance, based on the literature review and the analysis of the model results, are summarized in table 2

5.5 Generating pseudo-absence data

Next, the pseudo-absence data was generated based on the 3-step approach presented in the background information section. In this case, the presence data consisted of all the grids that had a frog count higher than 0 and the absence data consisted of all the grids having a frog count of 0. Each of the grids with a frog count of 0 was considered to be a pseudo-absence point.

A simple k-means clustering algorithm was then trained on the presence dataset, creating clusters based on how environmentally similar each datapoint was to the others. Then the absence data was fed to the model.

It was decided upon that the number of pseudo-absences would be equal to the number of frog presences. This meant that, if the amount of presences was k, only the k most dissimilar datapoints (i.e. the datapoints that were removed the furthest from the centers of the clusters), would be added to the final dataset that was fed to the model.

The last step, of removing the data that was geographically too close, was not necessary as the division of the data into grids already ensured that the pseudo-absence points were sufficiently far removed from the presence points.

5.6 Model training

Finally, the eventual dataset, which now only contained the relevant features and unbiased datapoints, was fed to the model. The data was randomly split into a 70/30 split, in which 70% was used to train the model and 30% was left out to validate the results.

The model itself was a Neural Network, which had the selected features as inputs and the frog count amount of a grid as output.

6 RESULTS

In order to measure model performance, three experiments were conducted. Each time, three areas were selected, one for each of the countries that this research is focused on (Australia, Costa Rica and South Africa). The size of the selected areas was increased with each experiment, leading to more data points (grids) to train the model on. This was done in order to find out what effect the increase of data would have on model performance.

For each experiment, the areas were chosen manually, by plotting the frog presence data in each of the countries and selecting areas which were dense with frog presence data. Unsurprisingly, most frog presence data was available around large cities (e.g. Sydney and Melbourne) and national parks (e.g. the Kruger National Park), so these areas were mostly selected.

Each area used in the experiments consisted of a name (to make it easier to reference them) and a bounding box, containing the coordinates of the area in the following form: (minimum longitude, minimum latitude, maximum longitude, maximum latitude).

For each experiment, each of the steps described in the method section was performed, after which the model was trained once with presence-only data (i.e. all grids with frog count > 0) and once with a combination of presence-only data and pseudo-absence data. This way, the impact of the pseudo-absences could also be determined.

As mentioned earlier, the data was split into a train and a test set. The performance of the model was represented using the R2 score the model achieved on the test set. The R2 score "represents the proportion of variance (of y) that has been explained by the independent variables in the model. It provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance" [18].

So, the higher the R2 score is, the more the variance in the target variable can be explained by the variances of the independent variables. If the R2 score is 100%, the variance in the target variable can be entirely explained by the variances in the independent variables, indicating a perfect correlation between the independent variables

Fig. 6. Frog presence - Sydney area

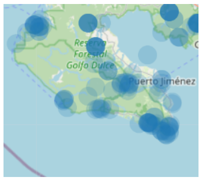

Fig. 7. Frog presence - Puerto Jimenez area

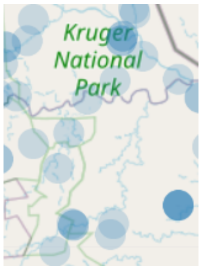

Fig. 8. Frog presence - Kruger National Park area

Table 3. Areas used in experiment 1

Area	Bounding box
Syndey	(150.5, -34.0, 151.0, -33.5)
Puerto Jimenez	(-83.5, 8.25, -83.0, 8.75)
Kruger National Park	(31.25, -24.25, 31.75, -23.75)

and the target variable. A higher R2 score therefore suggests better model performance on unseen samples.

6.1 Experiment 1: 100 grids per area

For the first experiment, three areas (see table 3 and figures 6-8) were chosen such that each of the areas could be split up into 100 (5.55 squared kilometer) grids. So there were 300 grids in total.

6.1.1 *Without pseudo-absences.* **R2-score: 0.001**

The first experiment resulted in an R2 score very close to 0, indicating a lack of correlation between the independent variables and the target variable. The cause of this was most likely the lack of data used to train the model on.

6.1.2 *With pseudo-absences.* **R2-score: 0.065**

Adding pseudo-absence data significantly increased the R2 score, it being over 60 times bigger than the R2 score that was achieved without using the pseudo-absence data. However, the accuracy was still very low, so it can be concluded that a sample count of 300 is not enough to train a Species Distribution Model for frogs.

Table 4. Areas used in experiment 2

Area	Bounding box
Melbourne	(144.5, -38.25, 145.5, -37.25)
San Jose	(-84.0, 9.5, -83.0, 10.5)
Kruger National Park	(31.0, -25.0, 32.0, -24.0)

Fig. 9. Frog presence - Melbourne area

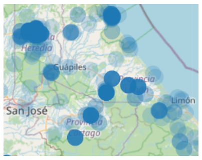

Fig. 10. Frog presence - San Jose area

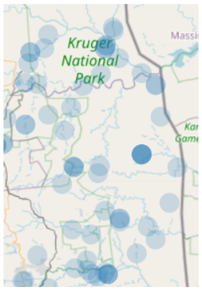

Fig. 11. Frog presence - Kruger National Park area (larger)

6.2 Experiment 2: 400 grids per area

For the second run, three larger areas were selected (see table 4 and figures 9-11), each of which big enough to be be split into 400 grids.

6.2.1 *Without pseudo-absences.* **R2-score: 0.264**

As can be seen, increasing the size of the areas led to a significant increase in the R2 score. This gives an indication that increasing the amount of data points does indeed lead to higher model accuracy.

6.2.2 *With pseudo-absences.* **R2-score: 0.365**

Even thought the impact was not as big as for the first experiment, the addition of pseudo-absences once again led to a significant increase in model performance, this time by over 10%, strengthening the belief that the use of pseudo-absences increases model performance.

Table 5. Areas used in experiment 3

Area	**Bounding box**
Sydney	(149.0, -34.75, 151.0, -32.75)
Most of Costa Rica	(-84.5, 8.5, -82.5, 10.5)
Kruger National Park	(30.5, -25.5, 32.5, -23.5)

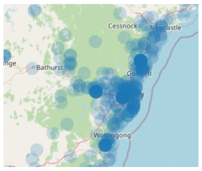

Fig. 12. Frog presence - Sydney area (larger)

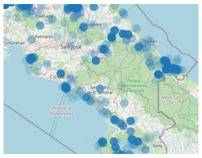

Fig. 13. Frog presence - Most of Costa Rica

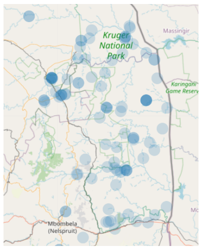

Fig. 14. Frog presence - Kruger National Park area (largest)

6.3 Experiment 3: 1600 grids per area

For the third and last experiment, three even larger areas were selected (see table 5 and figures 12-14), such that each of the areas could be split up into 1600 equal-sized grids.

6.3.1 Without pseudo-absences. **R2-score: 0.310**

There was another significant increase in model performance when increasing the amount of data used for the third time. Interesting was that the model did not perform better than it did during the previous run with pseudo-absence data.

6.3.2 With pseudo-absences. **R2-score: 0.546** For the last experiment, the inclusion of pseudo-absence data once again caused a significant increase in model performance. From this, it can be included that both increasing the amount of data, as well as using pseudo-absences, both consistently increase model accuracy.

7 CONCLUSION

This research tried to answer two questions. The first question was what predictor variables would be accurate for predicting frog distributions. This question was answered through a review of the existing literature, combined with an analysis of actual model performance. The predictor variables that were deemed to give the best indication of the size of a frog population are summarized in table 2.

The second question was whether a model could be built to accurately predict frog distributions, using the predictor variables. The results of the experiments conducted showed that this should indeed be possible. Even though the model could not be trained on larger areas of data due to limited computing resources, it was proved that the accuracy of the model significantly increased each time the amount of data fed to the model increased. This indicates that it should indeed be possible to accurately predict the size of frog populations, given that there are enough computing resources and data available.

Lastly, the results also showed that the use of pseudo-absence data did indeed increase model performance. From this it can be concluded that the use of pseudo-absence data is effective for predicting the size of frog populations when proper absence data is not available.

8 DISCUSSION

The main limitation for this research was the lack of computing resources, combined with the large nature of the data used (e.g. satellite data). This combination was the cause that only small subsets of the entire regions could be used for training the model.

The research has, however, proved that model accuracy increased significantly each time the amount of data fed to the model increased. The results of the research could therefore be improved upon in the future by using more computing resources.

Another potential improvement would be to use more datasets. For this research, only 5 datasets were used, but the Microsoft Planetary Portal offers many other datasets which could be used. Also, data could be gathered from other resources.

Improvement could also be made by distinguishing between frog species, as this research was focused on predicting frog distributions in general.

Also, for this research, absence data was not available and therefore pseudo-absence data was used instead. By training the model using actual absence data, it might become more accurate.

Lastly, this research focused on only three countries. In the future, a model could be trained using many other countries (e.g. countries from Europe could be included), which would make it more transferable across the entire world.

REFERENCES

[1] United States Naval Academy. 2019. Approximate Metric Equivalents for Degrees, Minutes, and Seconds. https://www.usna.edu/Users/oceano/pguth/md_help/html/approx_equivalents.htm

[2] The Editors of Encyclopaedia Britannica. 1998. indicator species. https://www.britannica.com/science/indicator-species

[3] Rosa M. Chefaoui and Jorge M. Lobo. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210, 4 (2008), 478–486. https://doi.org/10.1016/j.ecolmodel.2007.08.010

[4] Eric Chivian and Aaron Bernstein. 2010. How Our Health Depends on Biodiversity. (2010).

[5] J. Duque-Lazo, H. Van Gils, T. A. Groen, and R. M. Navarro-Cerrillo. 2016. Transferability of species distribution models: The case of Phytophthora cinnamomi in Southwest Spain and Southwest Australia. *Ecological Modelling* 320 (2016), 62–70. https://doi.org/10.1016/j.ecolmodel.2015.09.019

[6] Jane Elith and John R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology,*

Evolution, and Systematics 40, 1 (2009), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

[7] EOS. 2022. NDVI. https://eos.com/make-an-analysis/ndvi/

[8] EOS. 2022. NDWI. https://eos.com/make-an-analysis/ndwi/

[9] EY. 2022. EY launches 2022 Better Working World Data Challenge to help tackle biodiversity loss.

[10] EY. 2022. Level 3: Frog Counting Tool. https://challenge.ey.com/challenges/level-3-frog-counting-tool

[11] Alyssa A. Gerick, Robin G. Munshaw, Wendy J. Palen, Stacey A. Combes, and Sacha M. O'Regan. 2014. Thermal physiology and species distribution models reveal climate vulnerability of temperate amphibians. *Journal of Biogeography* 41, 4 (2014), 713–723. https://doi.org/10.1111/jbi.12261

[12] Nadia Shaker Habib, Omar Kamal Abu Maghasib, Ahmed Rashad Al-Ghazali, Bassem S. Abu-Nasser, and Samy S. Abu-Naser. 2020. Presence of Amphibian Species Prediction Using Features Obtained from GIS and Satellite Images. *International Journal of Academic and Applied Research (IJAAR)* 4, 11 (2020), 13–22.

[13] Microsoft. 2022. Data Catalog | Planetary Computer. https://planetarycomputer.microsoft.com/catalog

[14] Jennifer Miller. 2010. Species Distribution Modeling. *Geography Compass* 4, 6 (2010), 490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

[15] Laura J. Pollock, Reid Tingley, William K. Morris, Nick Golding, Robert B. O'Hara, Kirsten M. Parris, Peter A. Vesk, and Michael A. McCarthy. 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (<scp>JSDM</scp>). *Methods in Ecology and Evolution* 5, 5 (2014), 397–406. https://doi.org/10.1111/2041-210x.12180

[16] Clémentine Préau, Audrey Trochet, Romain Bertrand, and Francis Isselin-Nondedeu. 2018. Modeling potential distributions of three European amphibian species comparing ENFA and MaxEnt. *Herpetological Conservation and Biology* 13 (2018).

[17] Dennis Rödder and Stefan Lötters. 2010. Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (Eleutherodactylus planirostris). *Naturwissenschaften* 97, 9 (2010), 781–796. https://doi.org/10.1007/s00114-010-0694-7

[18] scikit learn. 2022. 3.3. Metrics and scoring: quantifying the quality of predictions. https://scikit-learn.org/stable/modules/model_evaluation.html#r2-score

[19] Senait D. Senay, Susan P. Worner, and Takayoshi Ikeda. 2013. Novel Three-Step Pseudo-Absence Selection Technique for Improved Species Distribution Modelling. *PLoS ONE* 8, 8 (2013), e71218. https://doi.org/10.1371/journal.pone.0071218