Species distribution modelling of the domesticated pig (*Sus domesticus*)

Jan Willem Wijnands (s2090554)

1 Introduction

The assignment within the course "Methods in Biodiversity Analysis" was about species distribution modelling. As input, the occurrence data of *Sus domesticus Erxleben, 1777* was used and retrieved from the GBIF database . This dataset consisted of 29 georeferenced datapoints concentrated in Europe with small occurrences in North America. Analyses of ancient mitochondrial DNA (mtDNA) from the domesticated pigs revealed that the domesticated pig was first transported by early famers form the Near East into Europe around 7,500 years before the present, also conforming zooarchaeological evidence for a single domestication origin of western Eurasian domestic pigs (Larson et al., 2007).

The global spread of these samples is shown in Figure 1. Climate data that was used during the species modelling was downloaded from the WorldClim database for the present and the future prediction for 2050 (HadGEM2-ES). The resolution of 5 minutes was chosen for both datasets.


Figure 1 Occurrences of data retrieved from GBIF database. Data is widespread over the globe. Little sample size.

2 Methods & Results

The species distribution modelling of *Sus domesticus* was carried out using the MAXENT software. For this analysis, the occurrence data was trimmed to the most compact sampled region (Europe) to construct the training set. This was done to avoid a sampling bias by the small spread of samples in North America. The species distribution model was cross-validated 8 times with 10.000 background sample points. Remaining parameters were set on "default" (Figure 2, 3). The model was used on the global climate data. Besides, a jackknife analysis was carried out to measure the variable importance.
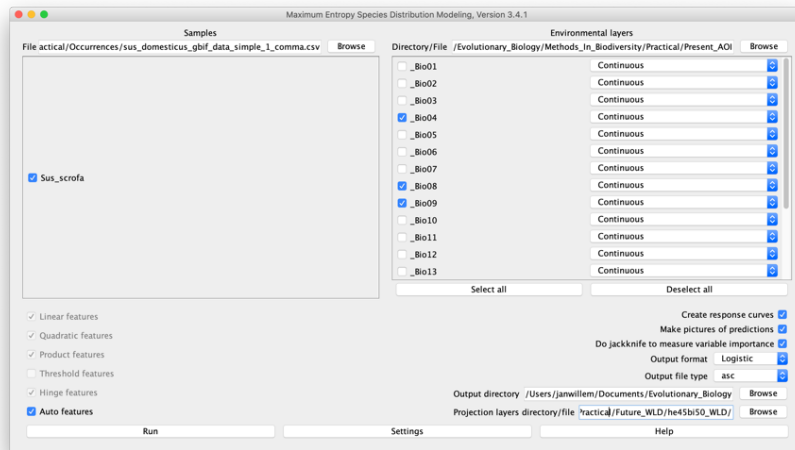
Figure 2 Main settings in MAXENT software.
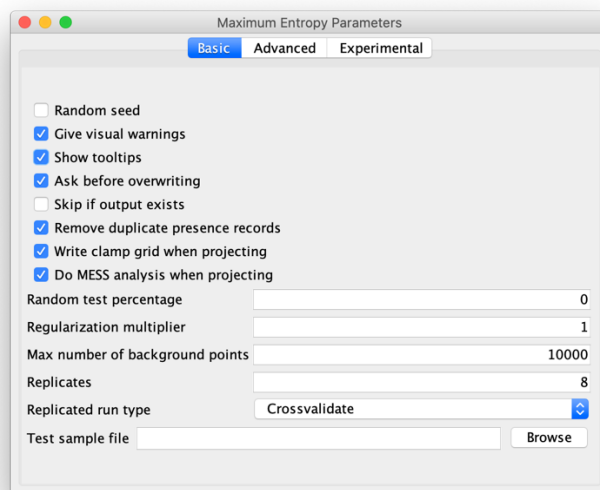

Figure 3 Settings used in MaxEnt

The bioclimatic variables input for the model, were encompassing the temperature seasonality, the mean temperature of the wettest quarter, mean temperature of the driest quarter and the precipitation seasonality (Table 1). These variables were selected because domestic pigs will not survive, when exposed to extreme wet or dry weather. Also, their food cannot be grown during these extreme periods.

**Table** Importance of the variables per bioclimatic variable within the model.

Bioclim. ID	Description	Contribution (%)
Bio4	Temperature seasonality	51.7
Bio8	Mean temperature of Wettest quarter	10
Bio9	Mean temperature of Driest quarter	27.7
Bio15	Precipitation seasonality	10.6

3.2 Output of MAXENT Software

The occurrence map of the sampled area is combined with the described bioclimatic variables (Bio4, Bio8, Bio9 and Bio15). This shows the suitability of the habitats for *Sus domesticus* through time, present (Figure 4) and the future (Figure 5). The future map is a prediction of the distribution of *Sus domesticus* until 2050 (RCP4.5).
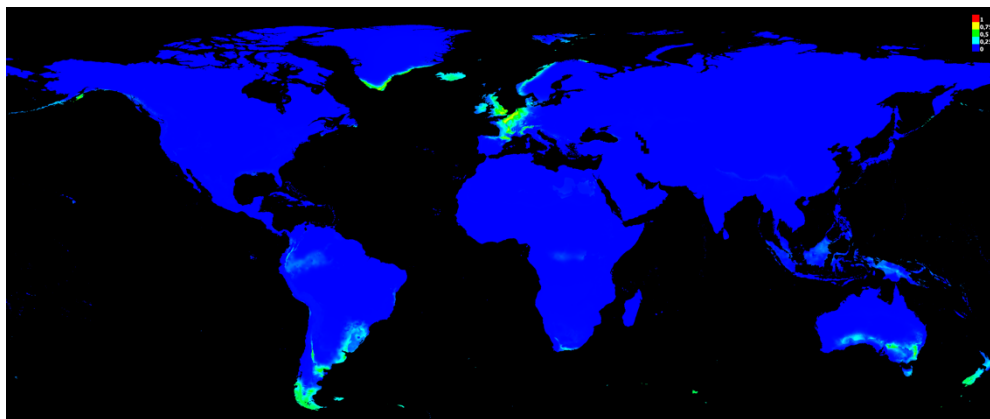

Figure 4 Present suitable habitat of *Sus domesticus*. Warmer colors describe suitable habitat for the species.
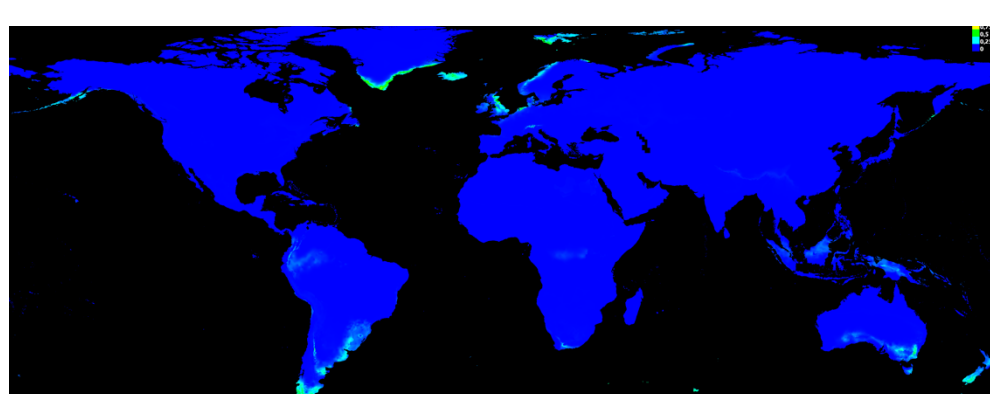
Figure 5 Habitat suitability in the future of *Sus domesticus*. Warmer colors describe suitable habitat for the species.

To test the reliability of the 8 validated models, a receiver operating characteristic (ROC) curve is constructed with use of the GBIF data. The ROC curve is unfortunately not fluent, because of the use of only 29 data points. However, the AUC value of training data equals 0.8, which means that created model is reliable. The model converges at a certain point because it finally can pick a point with matching AUC value. That is the difference between a species with over a few thousand of datapoints. Around 0.6 specificity (X-axis) there is one point that finish up the model (Figure 6). Because of less datapoints, this dataset was interesting to look into.
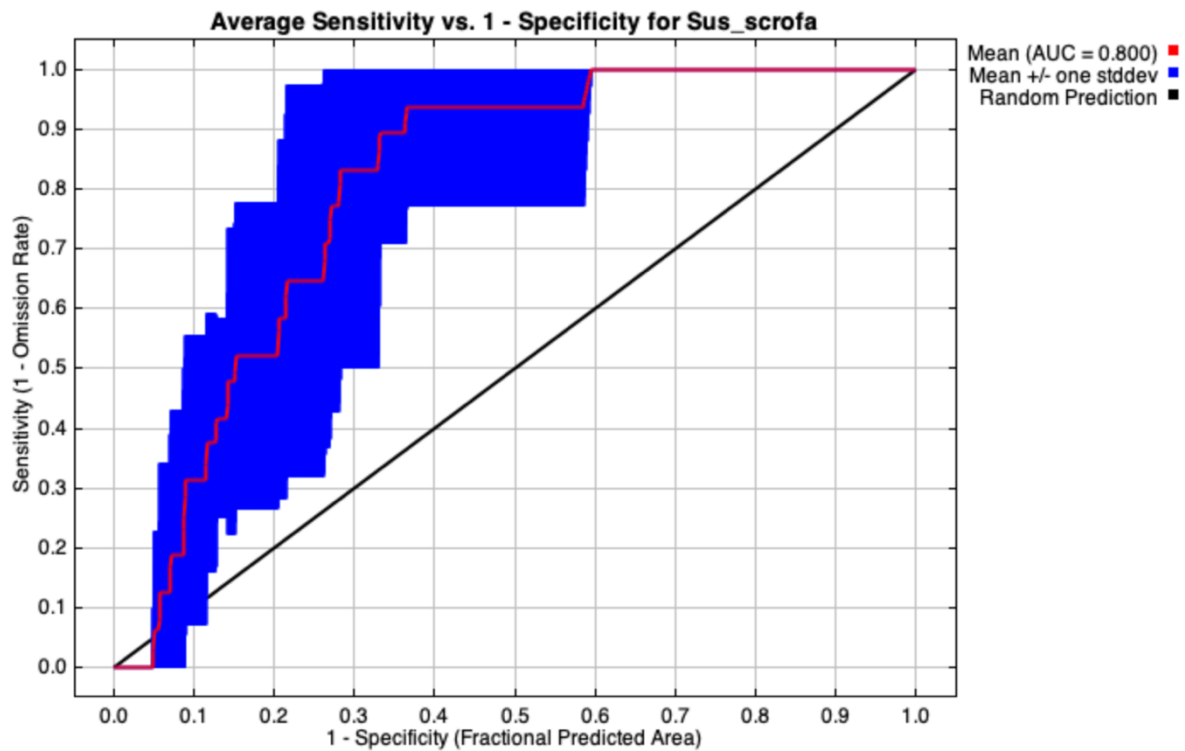
Figure 6 Average sensitivity graph. AUC= 0.8, which makes the model reliable. AUC>=0.7 means reliable/good.

3.2 Response to future scenario

In the last step during the analysis, the occurrence data from GBIF and the results from MaxEnt were combined to create a map. This map shows the suitability of different habitats for *Sus domesticus* populations (Figure 7). The map is cropped to the most dense occurrence data and focused on the difference in climate conditions.
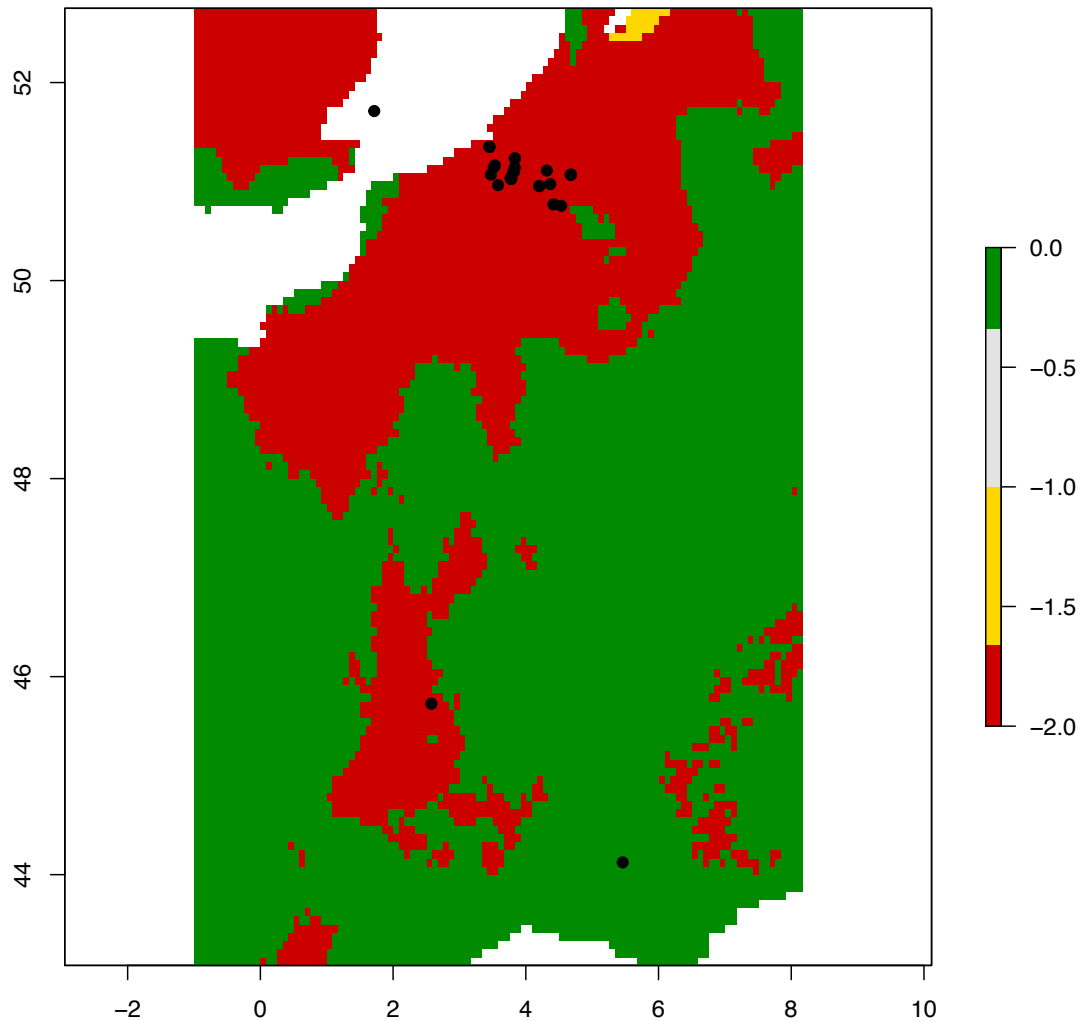
Figure 7 Difference in present and future in the suitable habitat of cropped Europe. Legend: green=area gained, yellow=remains suitable, red=lost area, black dots=sampling locations of data.

4 Interpretation of the results

The map shows the most losses of habitat, at RCP4.5, are in Europe. Unfortunately, it is not clear if this is completely true. We have to take into account the small sampling dataset in this conclusion. Mainly because of rising of the temperature a big loss of habitat will arise in the coming decades. This will have a major impact on the domestication of animals within Europe.

The AUC of this data can be defined as good. But we have to account for the use of data that only contained georeferenced data, which was a small fraction of the total amount of data on GBIF. Unfortunately, the model is not a reliable representation of the present and future SDM of *Sus domesticus* because of the use of only 29 sampling points. However, we were surprised about the 'positive' outcomes of the ROC curve.

5 References

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J.-D., Edwards, C.J., Schlumbaum, A., Dinu, A., Bescu, A., Dolman, G., Tagliacozzo, A., Manaseryan, N., Miracle, P., Wijngaarden-Bakker, L.V., Masseti, M., Bradley, D.G., Cooper, A., 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc. Natl. Acad. Sci. 104, 15276–15281. https://doi.org/10.1073/pnas.0703411104

GBIF.org (6th December 2018) GBIF Occurrence Download https://doi.org/10.15468/dl.htadl3

http://www.worldclim.org/version1