**Predicting future distributions of *Hordeum vulgare* using climate-based ecological niche modelling**

Name: Vincent Peelen

Student number: 1544357

Date: 12-12-2018

**Introduction**

*H. vulgare* is a cereal crop species belonging to the family Poaceae. While the distribution of its closely-related wild subspecies *H. spontaneum* (Middleton et al. 2014) is centered around the Middle East (figure 1) and is thought to have originally been domesticated in that area approximately 10000 years ago (Badr et al. 2000), the plant currently has a worldwide distribution according to occurrence data from GBIF (see references)*.* The present distribution could be attributed to its value as a crop species. However, climate change might cause this distribution to shift. Models utilizing bioclimatic variables can be used to predict this shift.



Figure 1: Distribution of *Hordeum spontaneum,* the wild subspecies of *Hordeum vulgare.* The rectangles indicate areas on the map where *H. spontaneum* occurs, the size of the red dot gives an estimation of the relative amount of occurrences. Obtained from https://www.gbif.org/occurrence/map?taxon\_key=2706075

**Methods**

To investigate the present and future habitat suitability for *H. vulgare,* a model had to be created that took into account the distribution of the crop species, presently known climate data and predicted changes in climate. The distribution data of *H. vulgare* includedwas obtained from GBIF with coordinates (53455 occurrences, see references for DOI). Present and future bioclimatic variables at 5 minute resolution were obtained from WorldClim, with the future variables using the climate model HadGEM2-AO (2070, RCP 4.5, pr).

RStudio (v1.1.442) was used to determine and prepare the bioclimatic predictor variables to be used in the model, using the packages *biomod2* (v3.37), *sp* (v1.3-1), *rgdal* (v1.3-6), and *raster* (v2.7-15).

The chosen predictors were Bio 4 (temperature seasonality), Bio 2 (mean diurnal range) and Bio10 (mean temperature of the warmest quarter) to account for vulnerability to more extreme temperatures, and Bio 12 (annual precipitation) and Bio 17 (precipitation of the driest quarter) to take increased droughts into account (O’Donnell et al. 2012). The other predictors were eliminated to avoid autocorrelation. RStudio was also used to produce graphs of the present and future potential distribution, as well as a graph depicting the change in potential distribution.

The species distribution model was created in Maxent (version 3.3.3k) using the bioclimatic predictor variables selected in RStudio. The bioclimatic variables for present conditions were used as the environmental layers. The random test percentage was set to zero, and duplicate presence records were culled. The maximum number of background points was set to 54500 to fit the number of presence records available in the data. In addition, response curves were produced to observe the effect of each variable on the species. The rest of the settings were left in their default state.

**Model output**

The model generated by MAXENT has an AUC of 0.831, which indicates that it has a high accuracy and reliability.

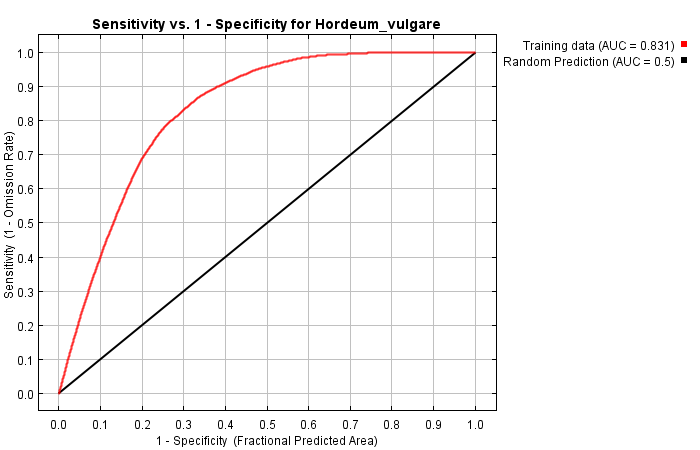


Figure 2: ROC curve of the model.

The variables Bio4 (temperature seasonality) and Bio10 (Mean temperature warmest quarter) contribute the most to the model, respectively 55% and 31.1%.

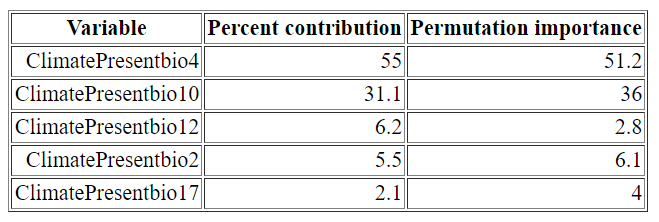


Table 1: Contribution and permutation importance for each variable.

Maxent produced the following logistic thresholds for this model:

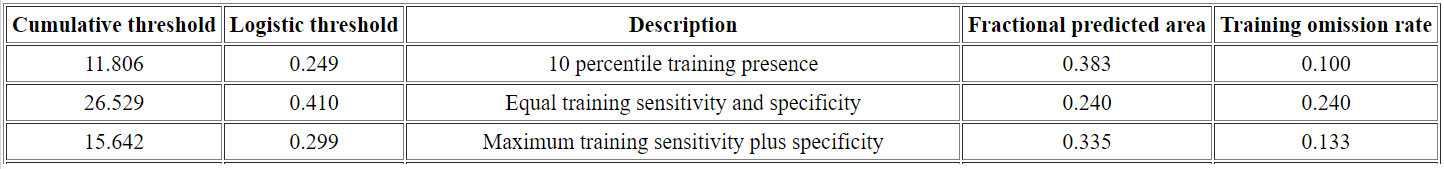
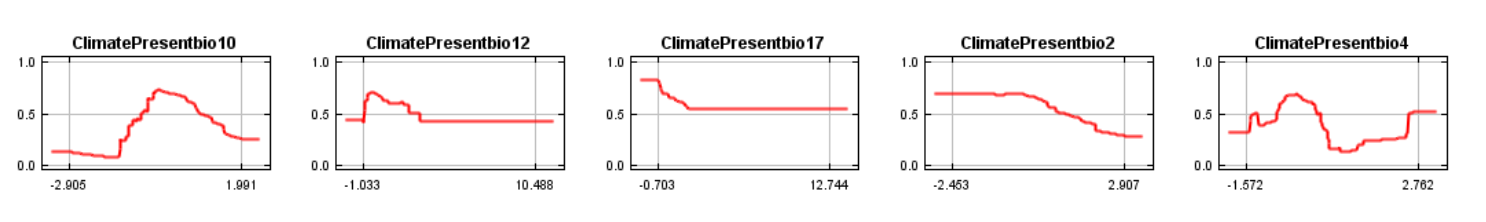


Table 2: Thresholds and omission rates.

The response curves show that *H. vulgare* is strongly affected by bio10 and bio4, confirming the percent contribution in table 1.



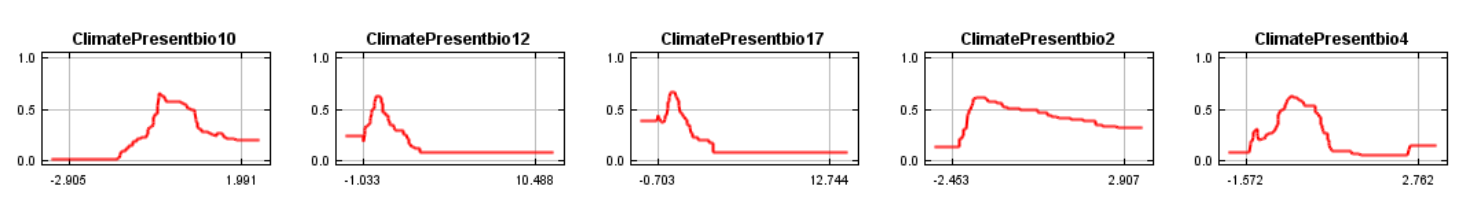


Figure 3: Response curves for the different climatic variables. The top row uses the climatic variables in the full model, whereas the bottom row uses a special model using only the listed variable.

MAXENT produced two graphs showcasing both the present and the projected distribution areas of *H. vulgare*, with slight differences in distribution between the two.

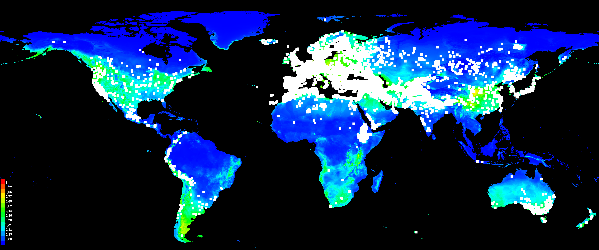
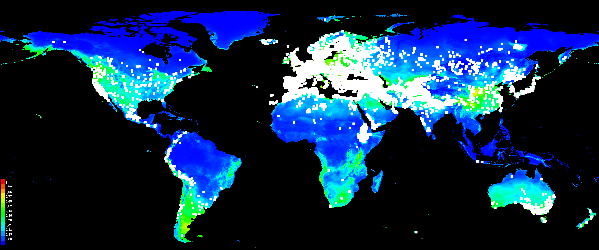
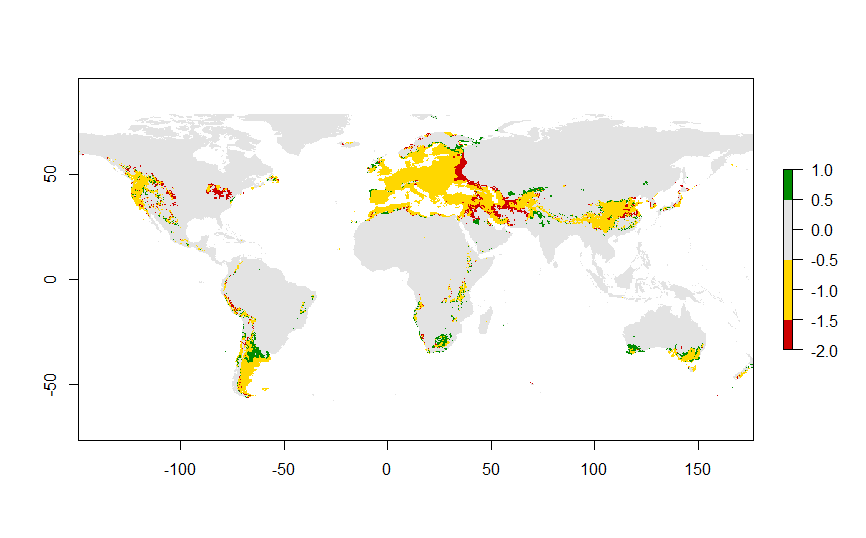
 

Figure 4 & 5: Top: Present suitable distribution area. Bottom: projected suitable distribution area. Occurrences are represented as white squares.

RStudio produced a graph showcasing the distribution area with the projected shifts highlighted.



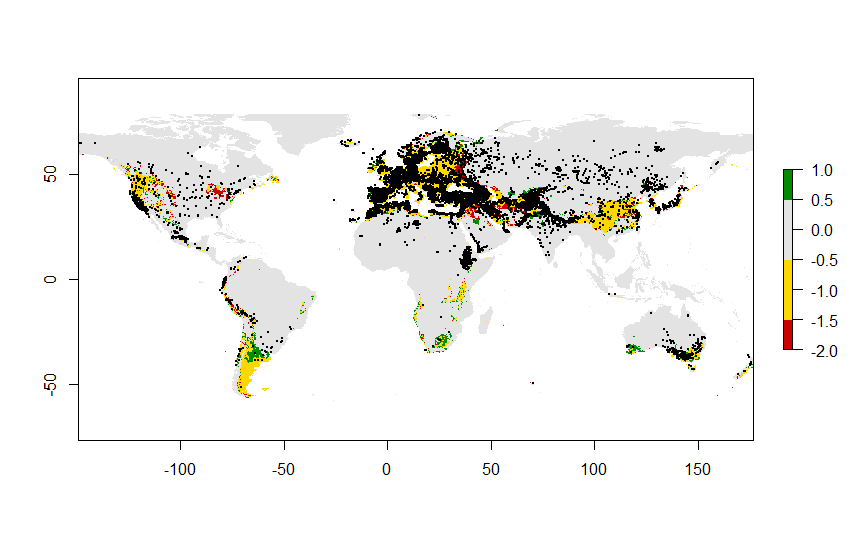


Figure 6 and 7: Gains and losses of suitable distribution area for *H. vulgare.* Distribution area that is not projected to change is marked in yellow. Distribution area growth is marked in green, and area loss is marked in red. Occurrences are marked with black dots on the bottom figure.

**Discussion**

The maps produced by RStudio and MAXENT indicate that the potential distribution area of *H. vulgare* may shift in the future. Small increases in area are found globally, with larger increases occurring in South America, South Africa and the southwestern corner of Australia. The climatic envelope matches occurrence data to a large degree, but not completely. Many ‘mismatched’ occurrences spread across North America, Russia and China, with remarkable clusters located in Ethiopia and northern China. At the same time, the southernmost part of South America has notably few occurrences considering the size of the area. This may indicate that the distribution of *H. vulgare* is not solely influenced climatic factors. For example, the large region of potentially suitable area in South America is largely mountainous, which could explain the lack of *H. vulgare* occurrences despite favourable climatic conditions.

Bio4 and Bio10, both temperature variables, make the most important contributions to the model, meaning that *H. vulgare* may be sensitive to changes in temperature. Neither of the precipitation variables made a strong impact on the model, which could indicate that the plant is drought-resistant. This makes sense when the distribution of its wildtype subspecies *H. spontaneum* is taken into account, as the frequency of its occurrences are most dense around the Middle East.

The model is incomplete, as it does not cover the entire spread of *H. vulgare* as indicated by the occurrence data. Adding different (a)biotic variables, such as elevation, soil type and biome type could help elaborate on the distribution.

**References**

Badr, A., Sch, R., Rabey, H. E., Effgen, S., Ibrahim, H. H., Pozzi, C., ... & Salamini, F. (2000). On the origin and domestication history of barley (Hordeum vulgare). Molecular Biology and Evolution, 17(4), 499-510.

GBIF.org (10 December 2018) GBIF Occurrence Download <https://doi.org/10.15468/dl.z1wf0k>

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., & Kilian, B. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. PLoS One, 9(3), e85761.

O’Donnell, M. S., & Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States. US Geological Survey Data Series, 691(10).