Species Distribution Modeling of *Hordeum vulgare*



Richard Frische s2096161

Methods in Biodiversity Analysis   
Leiden University, Winter 2019/20

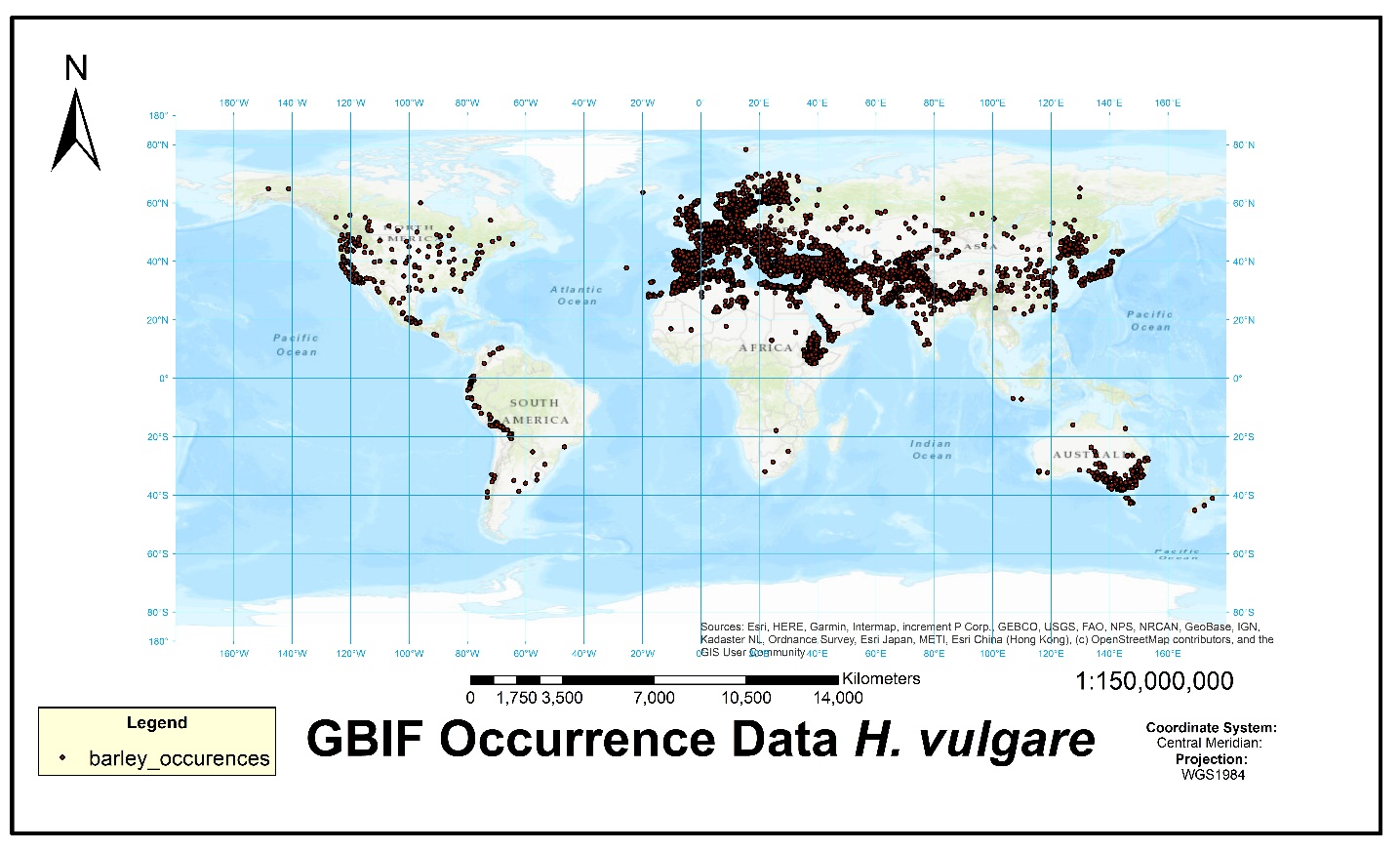
## Introduction

*Hordeum vulgare* (L.) is a major grain crop with ca. 140 million metric tons being produced in 2004.1 It is culturally significant for its role in malting and the production of alcoholic beverages such as beer and whiskey. Taxonomically it is placed within the family of Poaceae and is estimated to have diverged from other well-known grains such as wheat and rye ca. 8.9MYA.2 It is thought to have been domesticated ca. 10,000 years ago in the area of the “Fertile Crescent”.3 The requirements of barley for optimum growth dictate a climate with average annual temperatures of around 16°C and rainfall of ca. 400mm.4 It is tolerant of high temperatures when humidity is low but is less tolerant to frost in certain stages and prefers conditions with rainfall concentrated in the winter months.5 This report aims to illuminate possible shifts in range for *Hordeum vulgare* in light of projected climate change in line with Representative Concentration Pathways. As it is dependent on the aforementioned climatic conditions common in much of the species native range, a potential shift and large increase in suitable areas to more extreme southern and northern latitudes than which are currently in its range are hypothesized. Species distribution modeling will be applied to model potential shifts in range for *H. vulgare* in projected climate scenarios in dependence on relevant climate variables.

## Methodology & Results

Preparing of species occurrence data:

Species occurrence Data was retrieved from GBIF by querying for human and machine observations of *Hordeum vulgare* including coordinate data.6 The tab delimited comma separated variable file delivered by this query was reduced to species, latitude and longitude columns. Distribution of occurrences was then inspected in ArcMap on a topographical map in WGS1984 projection for biases and erroneous point data (Fig.1). Several points were removed from “seaborne” locations. The final dataset contained a total of 39,289 observations of which some are spatial duplicates resulting from multiple observations through time at the same location, these are later removed with the option “remove duplicate presence records” in the settings of MaxEnt. Certain sampling strategies are visible here with fixed patterns along grids visible in some countries along with a bias toward certain areas being most likely more well documented for occurrences of *H. vulgare* than other.

  
Fig. 1: Occurrence data for H. vulgare from GBIF plotted on topographical map. The circumglobal spread of this species is visible with the most occurrences being recorded throughout North Africa, Europe and across the Middle East and the Indian Subcontinent.

The occurrence data accurately reproduces the expected extent of barley as a domesticated crop species and additionally supports its character as an alpha-euhemerobe to polyhemerobe, that is its dependence on habitat with anthropogenic influences7 as well as it ecology and preference for seasonality in precipitation and tolerance to certain temperature ranges.5

Species distribution modeling:

Global climate data for multiple representative concentration pathways (2.6, 4.5, 6.0 & 8.0 respectively) and present climate data for 19 climate variables (see X) were retrieved at a spatial resolution of 5 arc-minutes from worldclim.org, version 1.4 was used.8 To generate a species distribution model for *H. vulgare* the maximum entropy approach will be used through use of MaxEnt software. On the basis of the obtained environmental and species occurrence data SDM was performed with MaxEnt.9 To prepare data for this process RStudio (Version 1.2.5019) is used to prepare an Area of Interest to train the SD model and read environmental files into raster data usable by MaxEnt. To select values of sufficient significance as a base for the model a pairwise autocorrelation test is first performed within the AOI. A table for all 19 bioclimatic variables is given with Pearson’s pairwise correlation values. The selection is made for the bioclimatic values where autocorrelation falls below 0.5, with a maximum negative correlation of -0.84 being achieved. The selection for the following test was narrowed down to the Bioclimatic variables 01, 02, 04, 07, 14, 15. After autocorrelation testing a variance inflation factor (VIF) test is performed. Initial runs with the variables suggest that either annual temperature range (Bio04) or seasonality (Bio07) must be removed due to collinearity of the variables, it was indicated by repeated VIF testing of these alternatives and consideration of ecological background (barley yields will be higher in more seasonal environments as growth phases are more steady)10 that removing the annual temperature range (Bio07) is most sensible as this gives VIF values of under 5 for most other variables which is a relatively conservative range. As a result the final selection of variables used in the model was the following: Bio01 Annual Mean Temperature, Bio 02 Mean Diurnal Range, Bio04 Temperature Seasonality, Bio14 Precipitation of Driest Month & Bio15 Precipitation Seasonality. Subsequent to this selection process MaxEnt can be run with the appropriate climate variables. We select to model for present data and RCPs 4.5, 6.0 & 8.5. Output is given in logistic form and clamping is activated ensuring that environmental variables are restricted to the range of values encountered during training of the model.11 A total of 5 replicates are run. MaxEnt produces a model with an AUC value of 0.724 (see Fig.2).

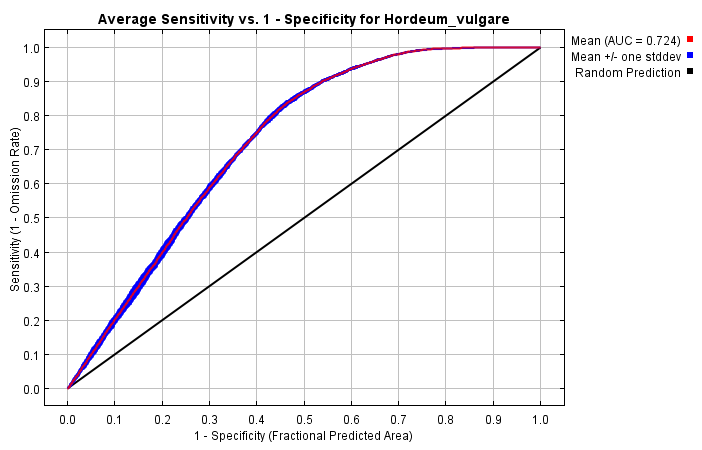
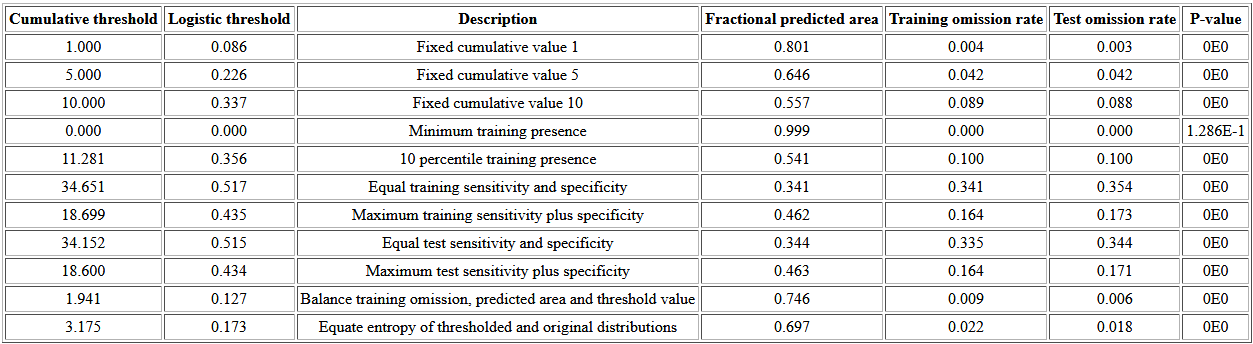


Fig. 2 Receiver Operating Curve. Averaged over replicated runs.

The average logistic threshold value for Maximum test sensitivity plus specificity across replicates is 0.431 (see Tab. 2).



Tab. 2 Predicted threshold values and omission rates.

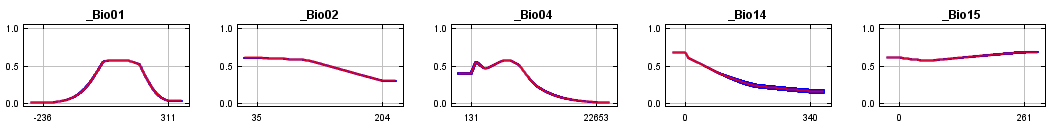
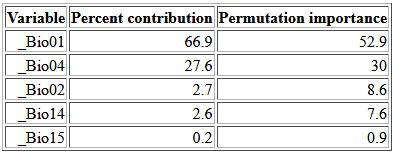
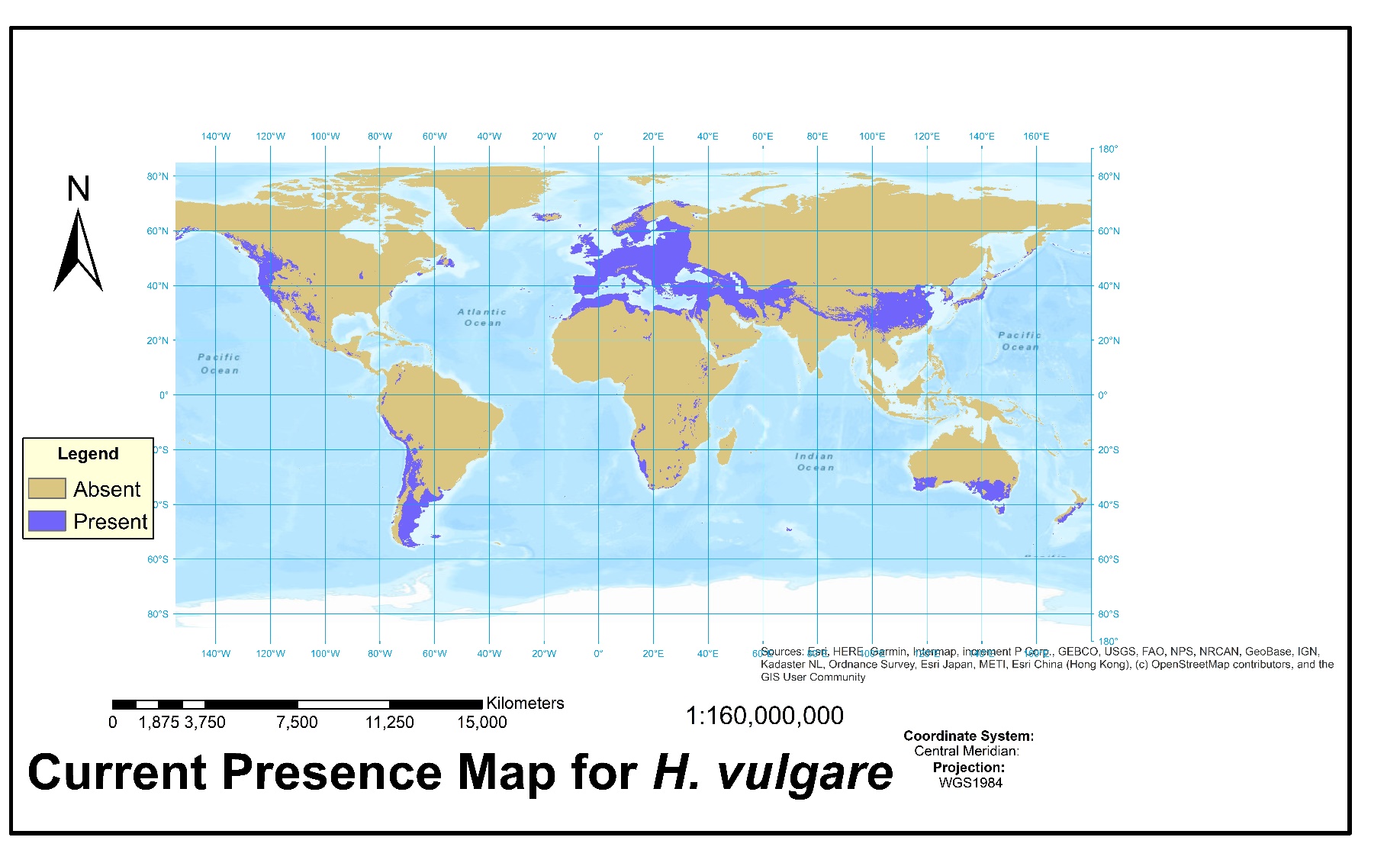


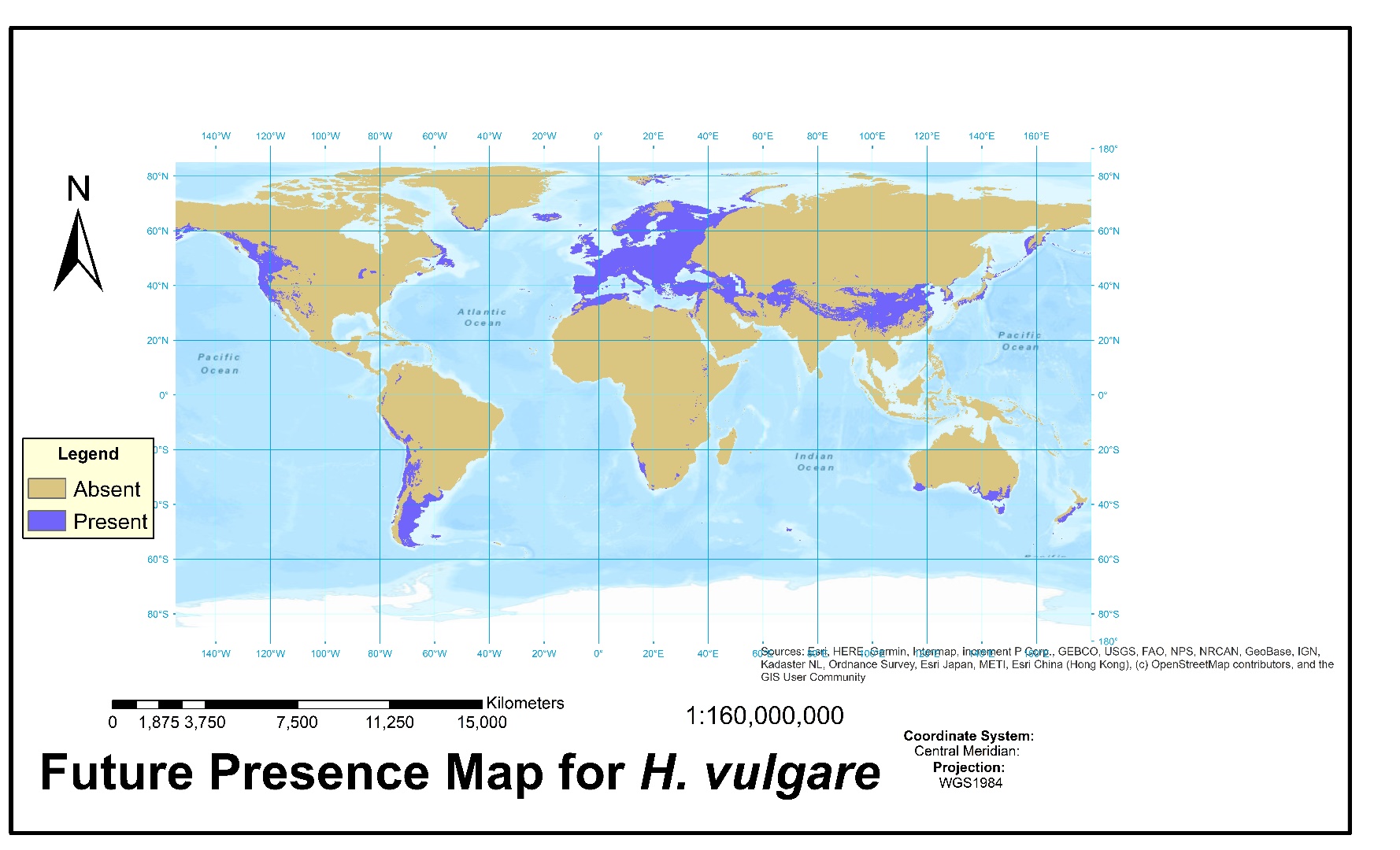
Fig. 3 Response curves indicating contribution of individual bioclimatic variables to the model.

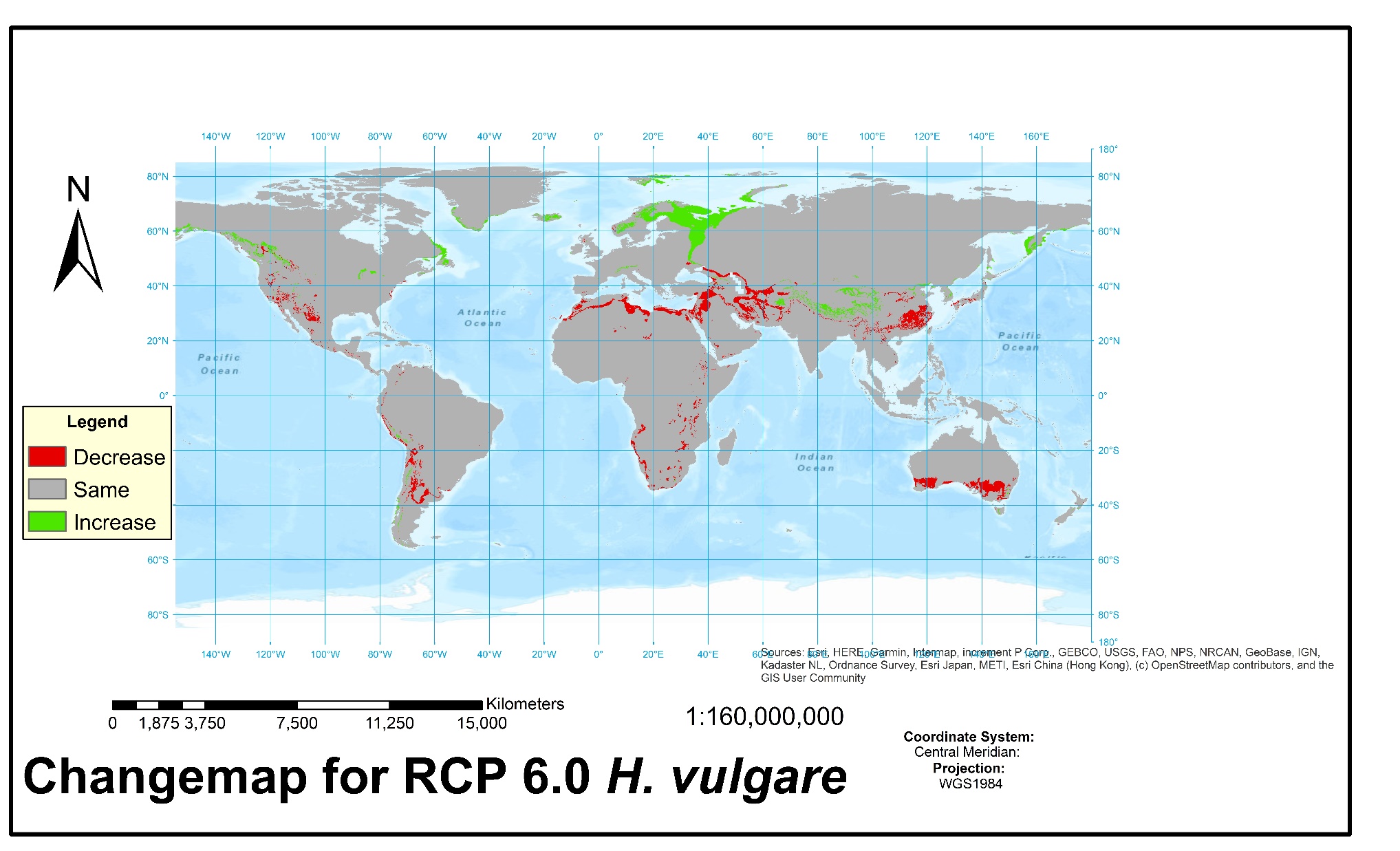


Tab. 3 Variable contribution table.

the large sampling area of globally widespread crops species such as *H. vulgare* and others, AUC Using the above threshold value changemaps are created in ARCGIS (Version 10.5.1).







Discussion:

The Changemap for RCP 6.0 given above indicates support for the hypothesis that *H. vulgare* will gain potential range in more northerly latitudes, this is visible especially 60° latitude north in Scandinavia and western areas of the Russian Federation as well as an increase in area predicted on the east and west coasts of North America above ca. 50°N and finally on the Kamchatka peninsula. Losses are predicted in barleys original home range of northwestern Africa and the Middle East where Mediterranean climates are projected to shift to conditions which are less ideal for this species. A similar trend in loss is visible across southern Australia where Barley is currently also present and grown in climates of Mediterranean character.5

Possible sampling bias may greatly influence the validity of SD modeling methods, as such it is of importance to carefully consider the generated models in biological and ecological context of the species to be modeled. In evaluating the produced model care must be taken to not rely heavily on singular measures of accuracy for the models such as the AUC. The achieved AUC of 0.724 is indicative of a reliable model in this case but must be treated with caution. With the large sampling area of globally widespread crops species such as *H. vulgare* and others, AUC may be increased by the inclusion of environmental values very different from those actually required by the species.12 The response curves (see Fig. 2) indicate in a general broad pattern the contributions of certain variables to the model that the test for variable importance supports (see Tab. 3). Bio01 (Annual Mean Temperature) contributes 66.9 percent and training AUC decreases 52.9 percent when omitted. None of the replicates indicated issues with clamping or training values being strongly dissimilar beyond some dissimilarity for Annual Mean Temperature in equatorial regions and Mean Diurnal Range in Greenland (see attached MaxEnt .html outputs GitHub). The threshold value selected for creating the change maps was Maximum test sensitivity plus specificity which is indicated for use in the current scenario by course material provided and literature on the subject.13 Overall this model indicates to a reliable extent and in sufficient sensitivity the projected species distribution for *H. vulgare* in a likely climate scenario, including larger datasets with verified reliable observations and employing a larger set of replicates than possible in the scope of this work may yield even more reliable models, that indicate trends shown here with improved accuracy.

Literature:

1. Gramene: A Resource for Comparative Grass Genomics-Barley Facts. Available at: https://archive.gramene.org/species/hordeum/barley\_facts.html. (Accessed: 10th December 2019)

2. Middleton, C. P. *et al.* 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**, e85761 (2014).

3. Badr, A. *et al.* On the Origin and Domestication History of Barley (Hordeum vulgare). *Mol. Biol. Evol.* **17**, 499–510 (2000).

4. Department of Agriculture Forestry and Fisheries Republic of South Africa. Brochure Barley. 2 (2009).

5. Australian Government- Department of Health and Ageing Office of the Gene Technology Regulator. The Biology of Hordeum vulgare L. (barley). 19–22 (2008).

6. GBIF.org. (2019). doi:10.15468/dl.tu3c3b

7. FloraWeb: Daten und Informationen zu Wildpflanzen und zur Vegetation Deutschlands. Available at: http://www.floraweb.de/pflanzenarten/oekologie.xsql?suchnr=10577&. (Accessed: 10th December 2019)

8. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

9. Phillips, S. J., Dudík, M. & Schapire, R. E. Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open\_source/maxent/. (2019).

10. Agriculture and Horticulture Development Board Cereals & Oilseeds. Barley Growth Guide 2018. (2018).

11. Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57 (2011).

12. Fourcade, Y., Engler, J. O., Rödder, D. & Secondi, J. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One* **9**, e97122 (2014).

13. Liu, C., White, M. & Newell, G. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789 (2013).

## Image sources:

https://bierverbond.nl/wp-content/uploads/2018/11/2018-nov-wat-is-lager-bier\_v001.jpg

https://d384u2mq2suvbq.cloudfront.net/public/spree/products/1376/jumbo/Whiskey\_Fragrance\_Oil.jpg?1486557222

http://www.homecookingadventure.com/images/recipes/sourdough-barley-bread.jpg

https://imbo.teknograd.no/s/OOPHs10