Species Distribution Modelling

Fagopyrum esculentum

Pip van Aalst
s1497952
Methods in Biodiversity Analysis
11-12-2019

http://commons.wikimedia.org/wiki/File:F 1

Inhoud

Introduction

The species that is selected for this project is the Fagopyrum esculentum or the common buckwheat. The buckwheat is a cultivated plant, that itself is not a cereal, but the seeds are used as cereal grains. This species is part of the Polygonaceae family, which is also referred to as the buckwheat or knotweed family. The plant has flowers that are grouped in clusters and has a fruit that is a triangular nut. The buckwheat usually grows 0,6-1,3 meters tall. It has a short taproot and lateral roots, which means that they can survive extreme drought, but this results in delayed growth. The branches turn reddish when they are matured. Short stamens and pistils are found in new flower forms that are adapted very well to self-pollination (Marshall, 1969). The buckwheat is an annual plant and cultivated in mostly the northern hemisphere.

It is thought that the buckwheat originated from the northwest corner of Yunnan province in China (Li & Yang., 1992; Ohnishi, 1995). It was cultivated for a millennium before it reached Europe in the 15th century (Hughes & Hensen, 1934). It was found that cultivation of the buckwheat in the Himalayas was limited to 500-2500 meters in altitude, above this altitude another species replaced buckwheat in its cropping patterns. A uniformity in allelic frequencies of isozyme loci was found in the buckwheat and is possibly linked to the recent cultivation of the species in this specific area. Another more plausible explanation is that the history of the cultivation of this species in the Himalayas is quite long, but because of repeated migration led to uniform allozymes (Ohnishi, 1988).

Joshi and Paroda (1991) found that the buckwheat is a short-duration crop (3-4 months) and that they require a moist and cool temperate climate. The buckwheat can't cope well with frost and is therefor usually grown at lower altitudes. The buckwheat also grows well on sandy and well-drained soils. When the moisture is low, the plant is more sensitive to higher temperatures.

Through Species Distribution Modelling the present occurrence data of the species is visualized and future prediction of the distribution of the data is made.

Research question

What is the present distribution of *Fagopyrum esculentum* and what will be the likely distribution in the future?

Hypothesis

It is expected that the *Fagopyrum esculentum* is present in areas with a cool and moist temperate climate. It is also expected that due to climate change the habitat of the species warms up. Therefor in the future scenarios the species is expected to move to more northern areas, where the temperature is cooler and also to higher altitudes for the same reason.

Methodology & results

Overview of steps

Collecting species occurrence data from GBIF → Collecting environmental data on the present and future scenarios from worldclim → Selecting environmental variables → Using MaxEnt to analyse the data → Analysing the models from MaxEnt → Making change maps in RStudio → Analysing the final results

Occurrence data

Occurrence data is extracted from GBIF and explored using ArcGIS. Most datapoint were present in areas with a temperate climate. There were some point present in oceans and rivers, these points were deleted from the data and not taken into account. Furthermore the occurrences seem to properly reflect the ecological niche.

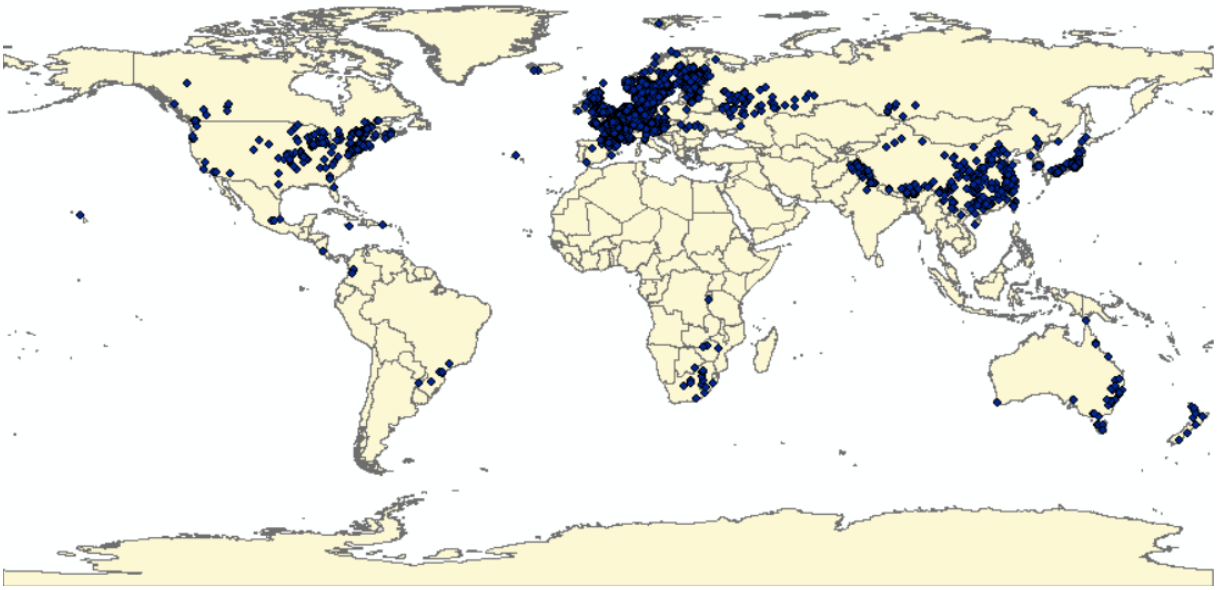


Figure 1: Occurrence data map of Fagopyrum esculentum made in ArcMap

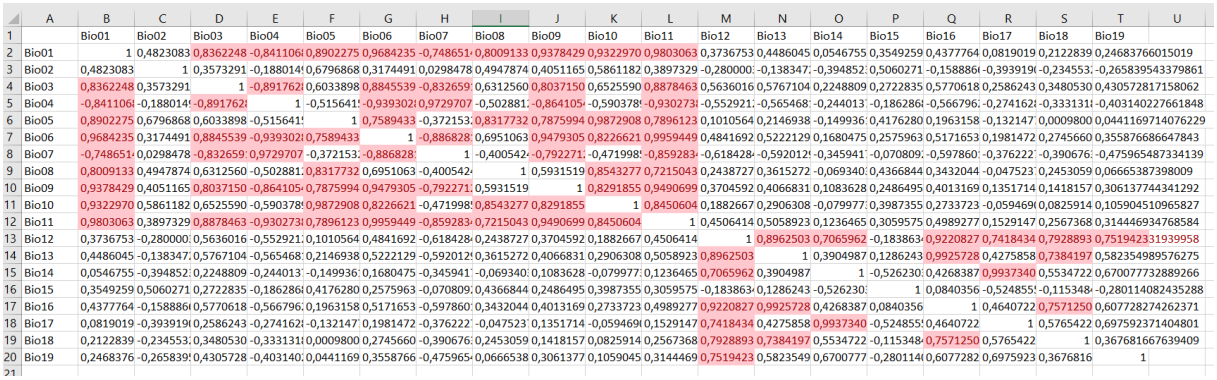
Environmental data

Buckwheat is a plant that grows best in a moist and cool temperate climate, but can't cope well with dry and warmer climates. Therefor the following environmental variables are chosen.

Bio-01: Annual Mean Temperature
Bio-04: Temperature Seasonality (later dismissed due to VIF value)
Bio-07: Temperature Annual Range
Bio-12: Annual Precipitation
Bio-15: Precipitation Seasonality
Bio-18: Precipitation Warmest Quarter

Autocorrelation

Environmental autocorrelation is measured by the Pearson's pairwise correlation test. Here the statistical correlations between two variables is measured.



	Bio01	Bio02	Bio03	Bio04	Bio05	Bio06	Bio07	Bio08	Bio09	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
Bio01	1	0,4823083	0,8362248	-0,841106	0,8902275	0,9684235	-0,748651	0,8009133	0,9378429	0,9322970	0,9803063	0,3736753	0,4486045	0,0546755	0,3549259	0,4377764	0,0819019	0,2122839	0,24683766015019
Bio02	0,4823083	1	0,3573291	-0,188014	0,6796868	0,3174491	0,0298478	0,4947874	0,4051165	0,5861182	0,3897329	-0,280000	-0,138347	-0,394852	0,5060271	-0,158886	-0,393919	-0,234553	-0,265839543379861
Bio03	0,8362248	0,3573291	1	-0,891762	0,6033898	0,8845539	-0,832659	0,6312560	0,8037150	0,6525590	0,8878463	0,5636016	0,5767104	0,2248809	0,2722835	0,5770618	0,2586243	0,3480530	0,430572817158062
Bio04	-0,841106	-0,188014	-0,891762	1	-0,515641	-0,939302	0,9729707	-0,502881	-0,864105	-0,590378	-0,930273	-0,552921	-0,565468	-0,244013	-0,186286	-0,566796	-0,274162	-0,333131	-0,403140227661848
Bio05	0,8902275	0,6796868	0,6033898	-0,515641	1	0,7589433	-0,372153	0,8317732	0,7875994	0,9872908	0,7896123	0,1010564	0,2146938	-0,149936	0,4176280	0,1963158	-0,132147	0,0009800	0,044116971407622 9
Bio06	0,9684235	0,3174491	0,8845539	-0,939302	0,7589433	1	-0,886828	0,6951063	0,9479305	0,8226621	0,9959449	0,4841692	0,5222129	0,1680475	0,2575963	0,5171653	0,1981472	0,2745660	0,355876686647843
Bio07	-0,748651	0,0298478	-0,832659	0,9729707	-0,372153	-0,886828	1	-0,400542	-0,792271	-0,471998	-0,859283	-0,618428	-0,592012	-0,345941	-0,070809	-0,597860	-0,376222	-0,390676	-0,475965487334139
Bio08	0,8009133	0,4947874	0,6312560	-0,502881	0,8317732	0,6951063	-0,400542	1	0,5931519	0,8543277	0,7215043	0,2438727	0,3615272	-0,069340	0,4366844	0,3432044	-0,047523	0,2453059	0,066653873980 09
Bio09	0,9378429	0,4051165	0,8037150	-0,864105	0,7875994	0,9479305	-0,792271	0,5931519	1	0,8291855	0,9490699	0,3704592	0,4066831	0,1083628	0,2486495	0,4013169	0,1351714	0,1418157	0,306137744341292
Bio10	0,9322970	0,5861182	0,6525590	-0,590378	0,9872908	0,8226621	-0,471998	0,8543277	0,8291855	1	0,8450604	0,1882667	0,2906308	-0,079977	0,3987355	0,2733723	-0,059469	0,0825914	0,105904510965827
Bio11	0,9803063	0,3897329	0,8878463	-0,930273	0,7896123	0,9959449	-0,859283	0,7215043	0,9490699	0,8450604	1	0,4506414	0,5058923	0,1236465	0,3059575	0,4989277	0,1529147	0,2567368	0,314446934768584
Bio12	0,3736753	-0,280000	0,5636016	-0,552921	0,1010564	0,4841692	-0,618428	0,2438727	0,3704592	0,1882667	0,4506414	1	0,8962503	0,7065962	-0,183863	0,9220827	0,7418434	0,7928893	0,751942331939958
Bio13	0,4486045	-0,138347	0,5767104	-0,565468	0,2146938	0,5222129	-0,592012	0,3615272	0,4066831	0,2906308	0,5058923	0,8962503	1	0,3904987	0,1286243	0,9925728	0,4275858	0,7384197	0,582354989576275
Bio14	0,0546755	-0,394852	0,2248809	-0,244013	-0,149936	0,1680475	-0,345941	-0,069340	0,1083628	-0,079977	0,1236465	0,7065962	0,3904987	1	-0,526230	0,4268387	0,9937340	0,5534722	0,670077732889266
Bio15	0,3549259	0,5060271	0,2722835	-0,186286	0,4176280	0,2575963	-0,070809	0,4366844	0,2486495	0,3987355	0,3059575	-0,183863	0,1286243	-0,526230	1	0,0840356	-0,524855	-0,115348	-0,280114082435288
Bio16	0,4377764	-0,158886	0,5770618	-0,566796	0,1963158	0,5171653	-0,597860	0,3432044	0,4013169	0,2733723	0,4989277	0,9220827	0,9925728	0,4268387	0,0840356	1	0,4640722	0,7571250	0,607728274262371
Bio17	0,0819019	-0,393919	0,2586243	-0,274162	-0,132147	0,1981472	-0,376222	-0,047523	0,1351714	-0,059469	0,1529147	0,7418434	0,4275858	0,9937340	-0,524855	0,4640722	1	0,5765422	0,697592371404801
Bio18	0,2122839	-0,234553	0,3480530	-0,333131	0,0009800	0,2745660	-0,390676	0,2453059	0,1418157	0,0825914	0,2567368	0,7928893	0,7384197	0,5534722	-0,115348	0,7571250	0,5765422	1	0,367681667639409
Bio19	0,2468376	-0,265839	0,4305728	-0,403140	0,0441169	0,3558766	-0,475965	0,0666538	0,3061377	0,1059045	0,3144469	0,7519423	0,5823549	0,6700777	-0,280114	0,6077282	0,6975923	0,3676816	1

Figure 2: Table of autocorrelation between the present environmental variables.

As seen in figure 1, the Bio-01 until Bio-11 are grouped together. The same can be seen for Bio-12 until Bio-19. This can be explained by the fact that the first 11 variables are all related to temperature and the last variables are related to precipitation.

Multicollinearity

First Bio-01, Bio-04, Bio-07, Bio-12, Bio-15, Bio-18 were the chosen environmental variables and they were tested on multicollinearity. The Variable Inflaction Factor (VIF) is the correlation of one variable in combination with a number of other variables.

	Variables	VIF
1	Bio01	5.088489
2	Bio04	45.949658
3	Bio07	34.444913
4	Bio12	4.173865
5	Bio15	1.463386
6	Bio18	2.868115

Figure 3: Results of VIF Test for the initial environmental variables. Bio-04 and Bio-07 have a high VIF, therefor one of these variables is deleted.

For the second run Bio-04 is taken out of the multicollinearity test and the VIF is calculated again.

	Variables	VIF
1	Bio01	2.888191
2	Bio07	3.394277
3	Bio12	4.163142
4	Bio15	1.367713
5	Bio18	2.854682

Figure 4: Results of VIF Test for the final environmental variables chosen. All the variables VIF are below 5.0 and therefor these are the final variables chosen.

Model settings

Species Distribution Modelling is done using MaxEnt and is validated by Area Under the Curve (AUC) and the Receiver Operator Curve (ROC). A threshold is set to get continues MaxEnt values. The AUC values range from 0 to 1, 0.5 indicates that the model accuracy is not better than random and 1.0 being a perfect model. After choosing a threshold the contribution of the variables in the model are calculated and later a jackknife test is used to measure the model performance using each variable individually. The MaxEnt is replicated 5 times.

Model output

Area Under the Curve (AUC)

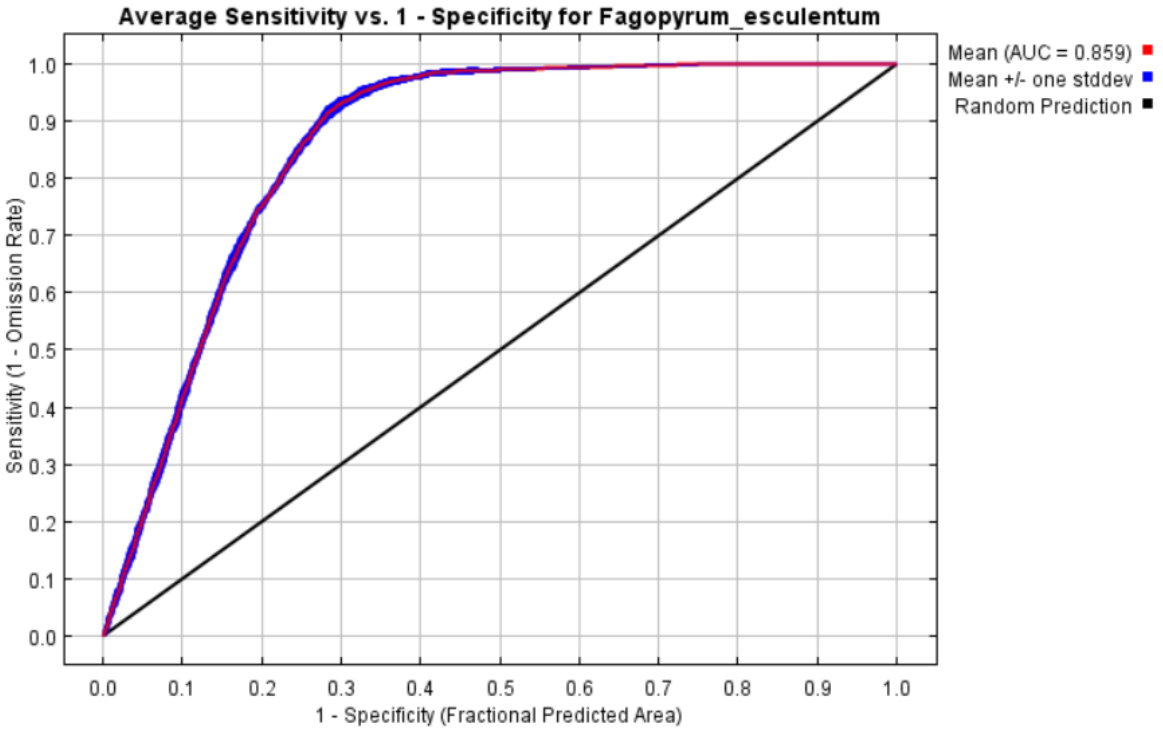


Figure 5: The average test AUC for the replicate runs is 0.859

As seen in figure 4 the AUC value is well over 0.5, which means that the accuracy of the model is pretty good.

Response Curves

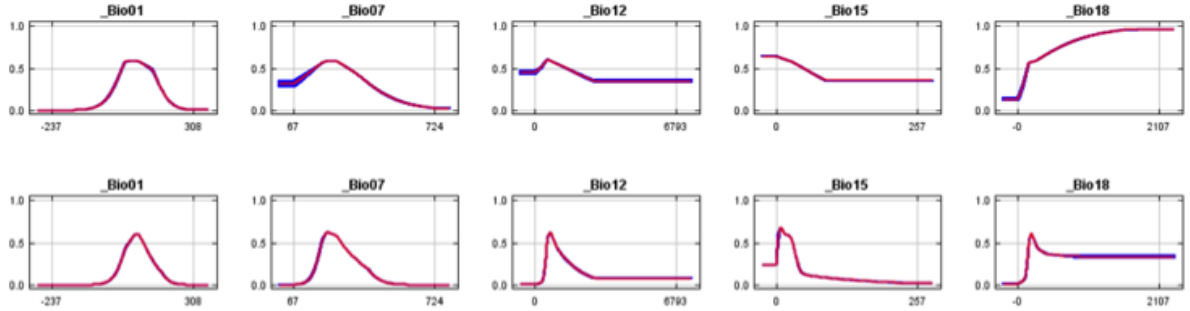


Figure 6: Response curves of the model. The figures in the top are the response curves per variable when all the other variables are also present in the model. The figures in the bottom are the response curves per variable when only that variable in present in the model.

Variable	Percent contribution	Permutation importance
_Bio01	50.5	65.7
_Bio12	23.1	2.5
_Bio07	13.1	15.1
_Bio15	9.1	5.1
_Bio18	4.1	11.5

Figure 7: Table with the relative contributions of the environmental variables.

Bio-01 or Annual Mean Temperature has the highest contribution to the model (figure 6) and this is also seen in figure 5. The graph of bio-01 are most similar to each other in figure 5. The same is for Bio-07 or Temperature Annual Range. This means the temperature has the highest influence on the model and the species distribution. Bio-12 or annual precipitation doesn't seem to matter to much for the species, when the other environmental variables are present in the model. When Bio-12 is analysed individually, the species is more sensitive to precipitation. The same is for Bio-15 or Precipitation Seasonality.

SDM projections

Climate data is extracted from the WorldClim website. All the data has a spatial resolution of 5 minutes. In the future predictions are for the year 2050 and with different greenhouse gas scenario's. There are four representative concentration pathways (RCP). The four RCPs are RCP2.6, RCP4.5, RCP6.0 and RCP8.5, which is in W/m^2.

Present World data

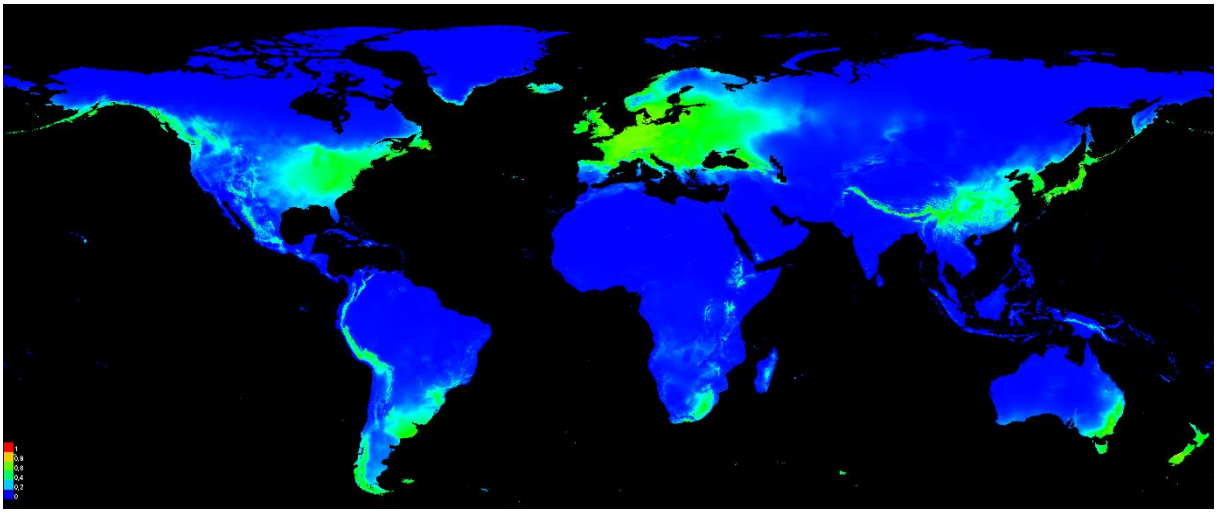


Figure 8: Present World data map of the present distribution of the species made in MaxEnt.

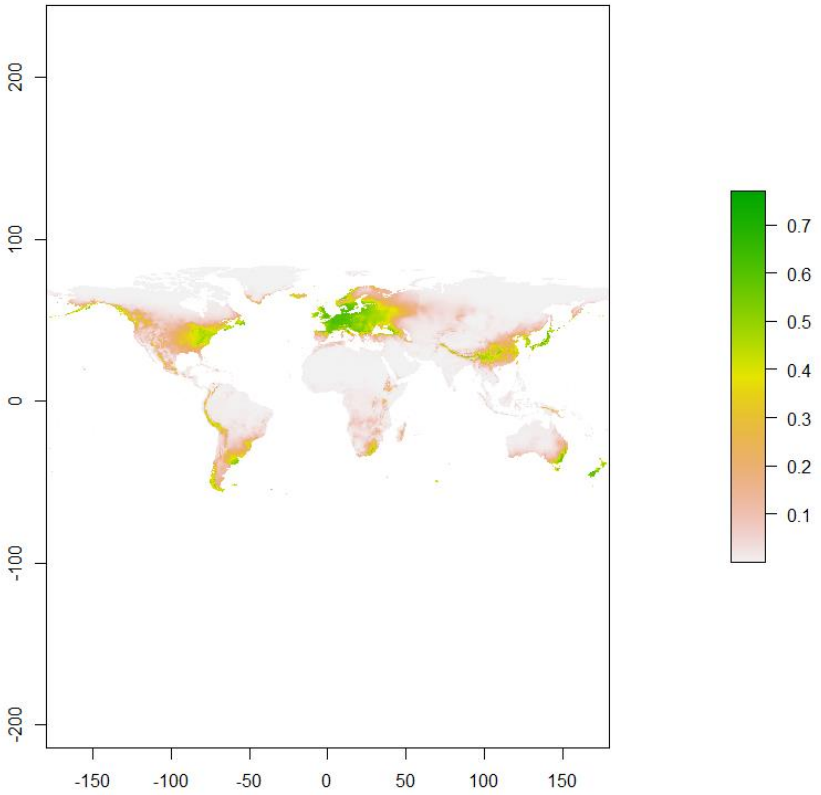


Figure 9: Present World data map of the present distribution of the species made in RStudio.

The species are mainly present in Europe, North America and China and Japan, all along the same axis. Less occurrences are in the southern hemisphere, but the occurrences present are also along the same axis. Most occurrences are in temperate and moist areas.

Future data

He26bi50

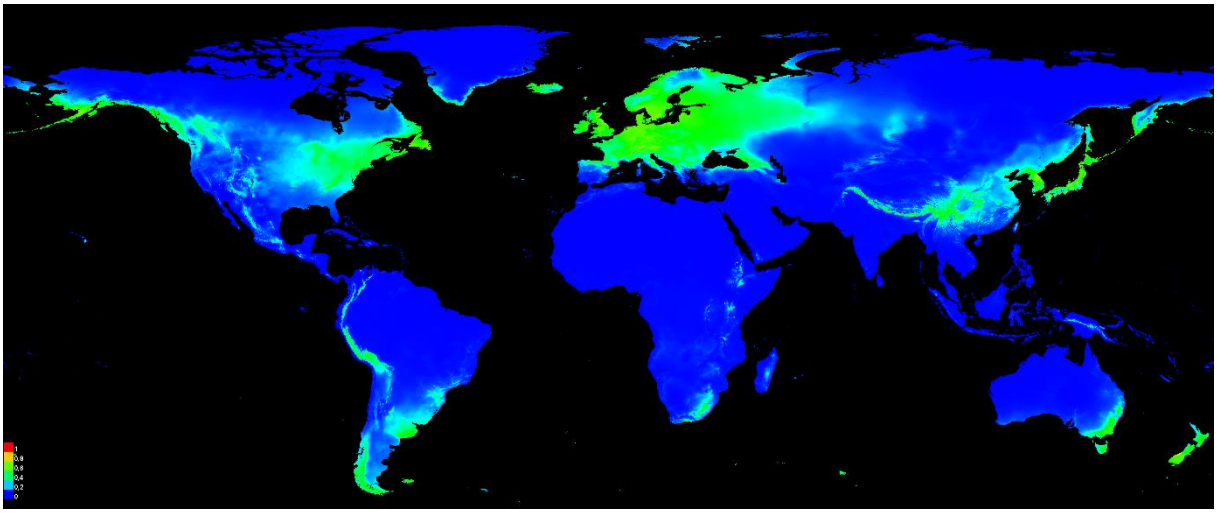


Figure 10: Future prediction of species distribution for RCP 2.6 made in MaxEnt.

This is the prediction of the species distribution in the year 2050 for RCP 2.6 W/m^2.

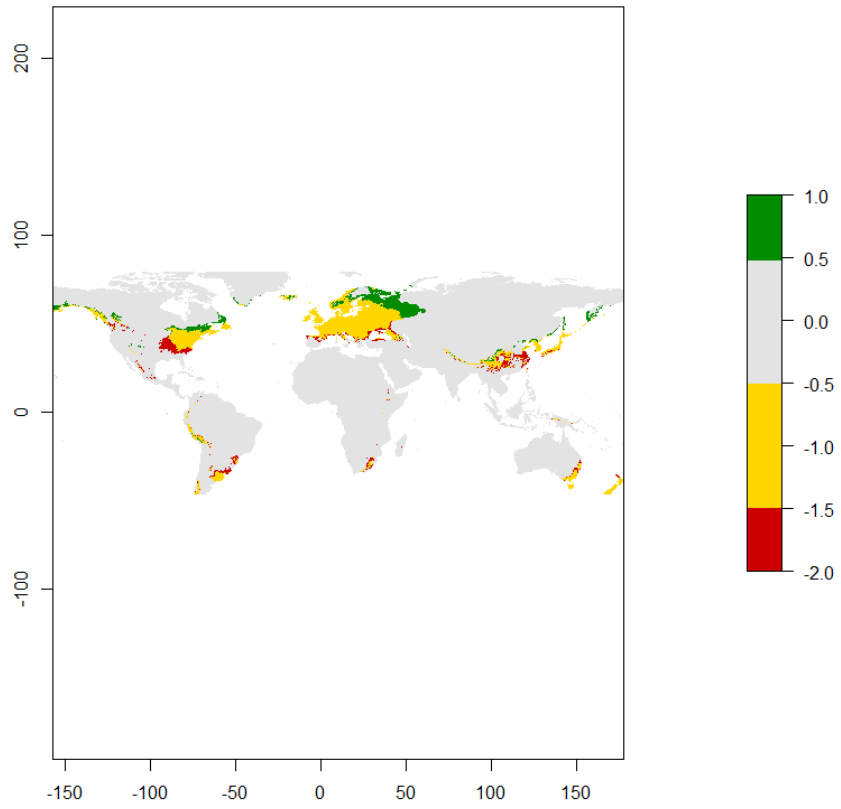


Figure 11: Change map of RCP 2.6 made in RStudio.

He45bi50

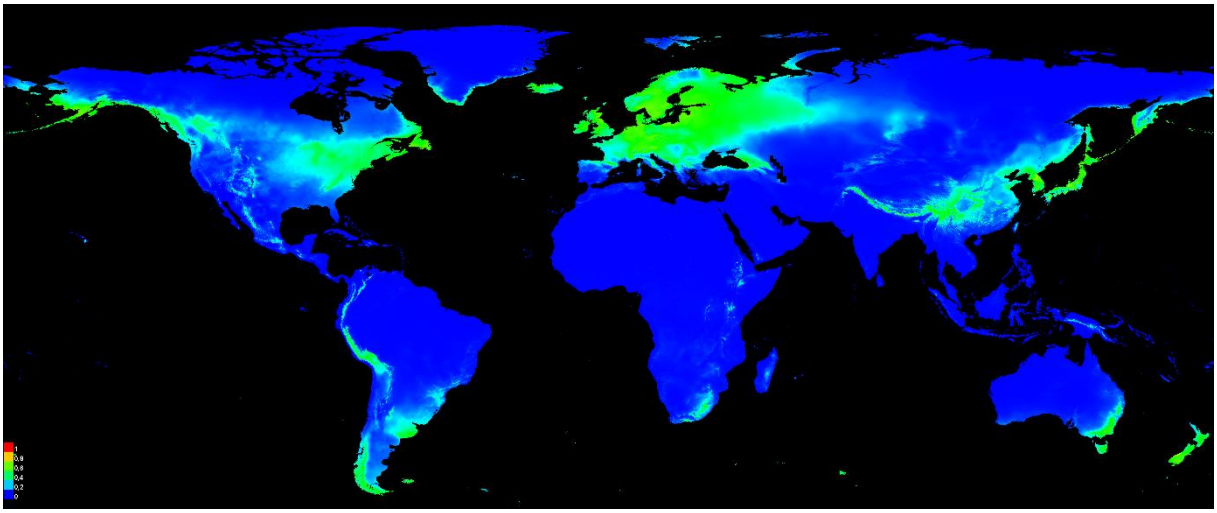


Figure 12: Future prediction of species distribution for RCP 4.5 made in MaxEnt.

This is the prediction of the species distribution in the year 2050 for RCP 4.5 W/m^2.

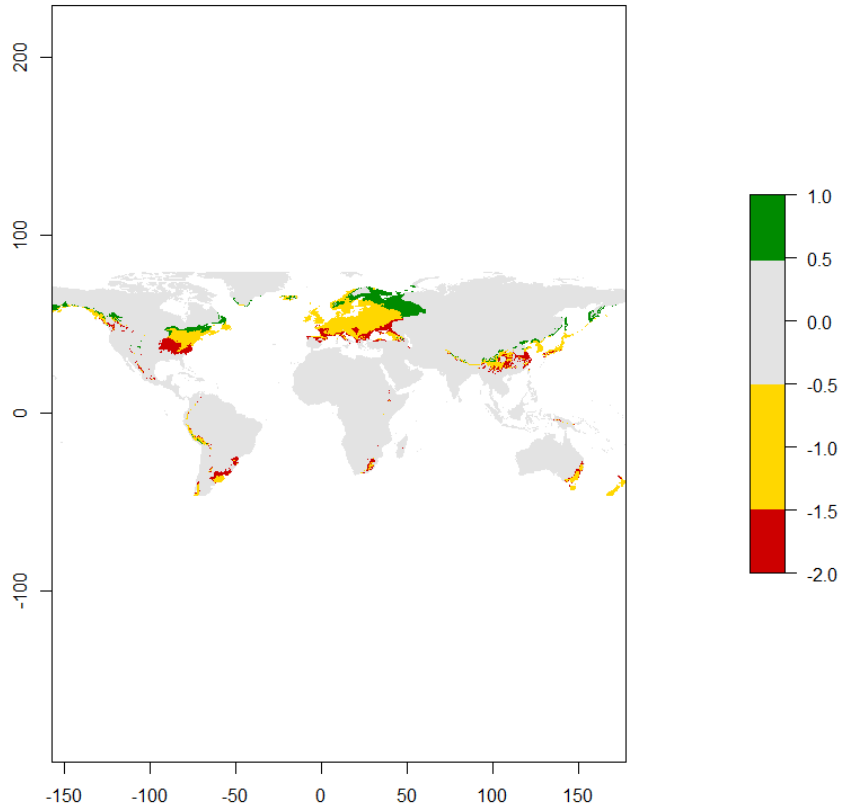


Figure 13: Change map of RCP 4.5 made in RStudio.

He60bi50

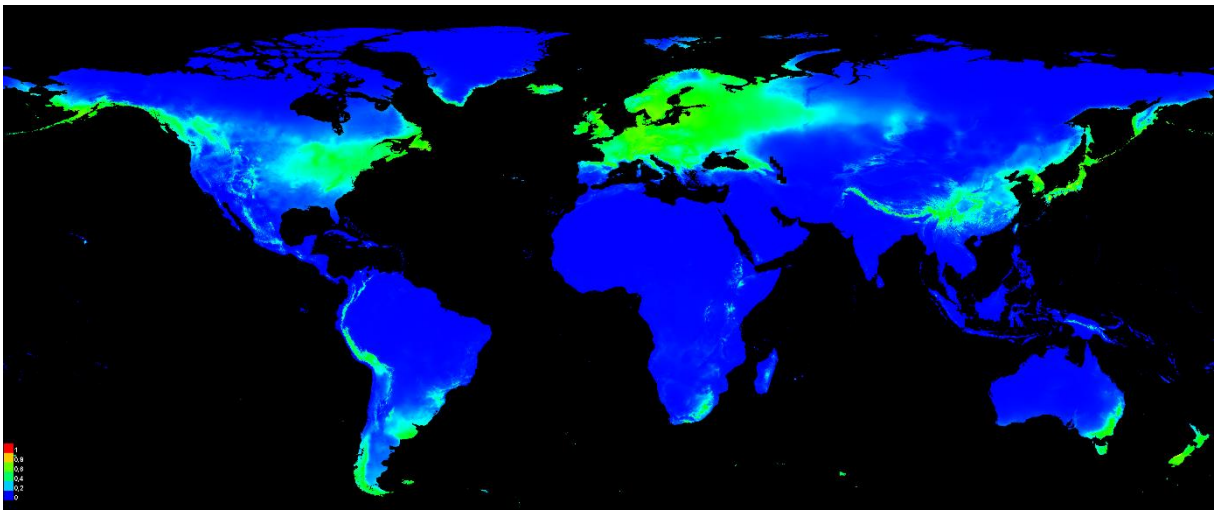


Figure 14: Future prediction of species distribution for RCP 6.0 made in MaxEnt.

This is the prediction of the species distribution in the year 2050 for RCP 6.0 W/m^2.

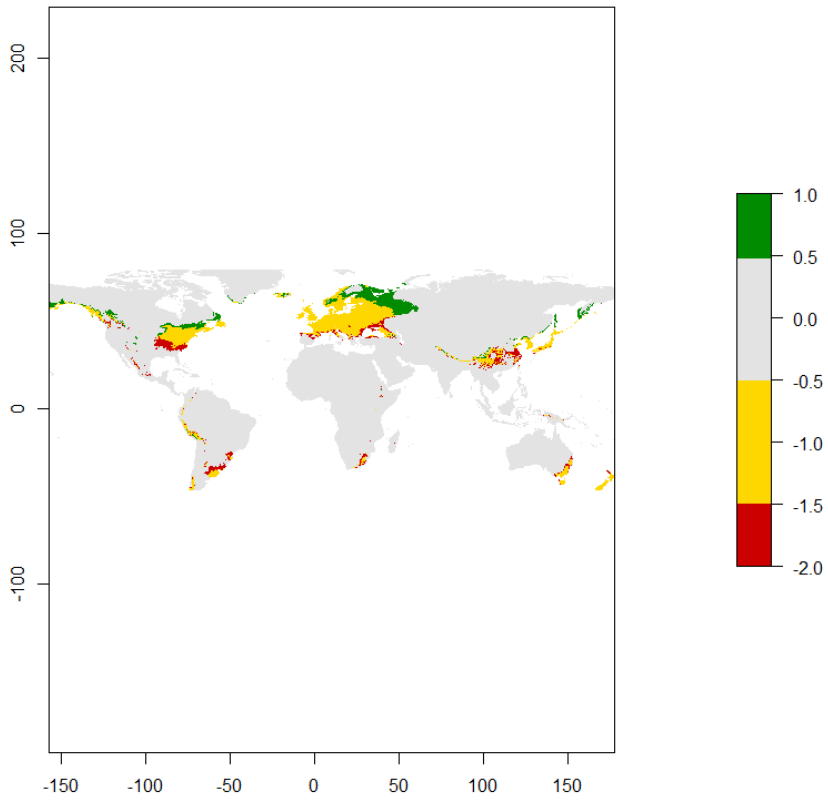


Figure 15: Change map of RCP 6.0 made in RStudio.

He85bi50

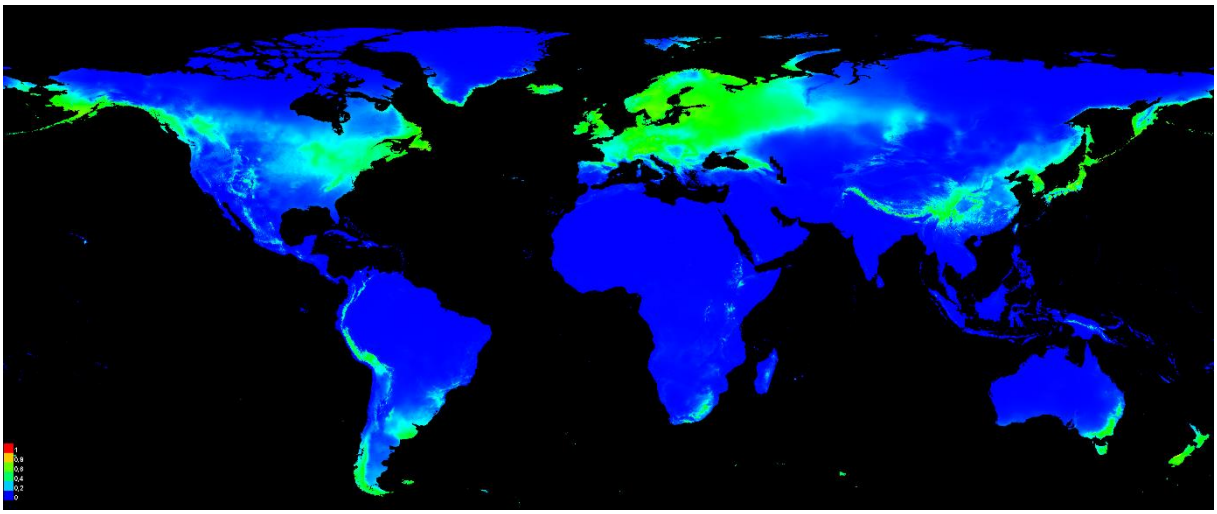


Figure 16: Future prediction of species distribution for RCP 8.5 made in MaxEnt.

This is the prediction of the species distribution in the year 2050 for RCP 8.5 W/m^2.

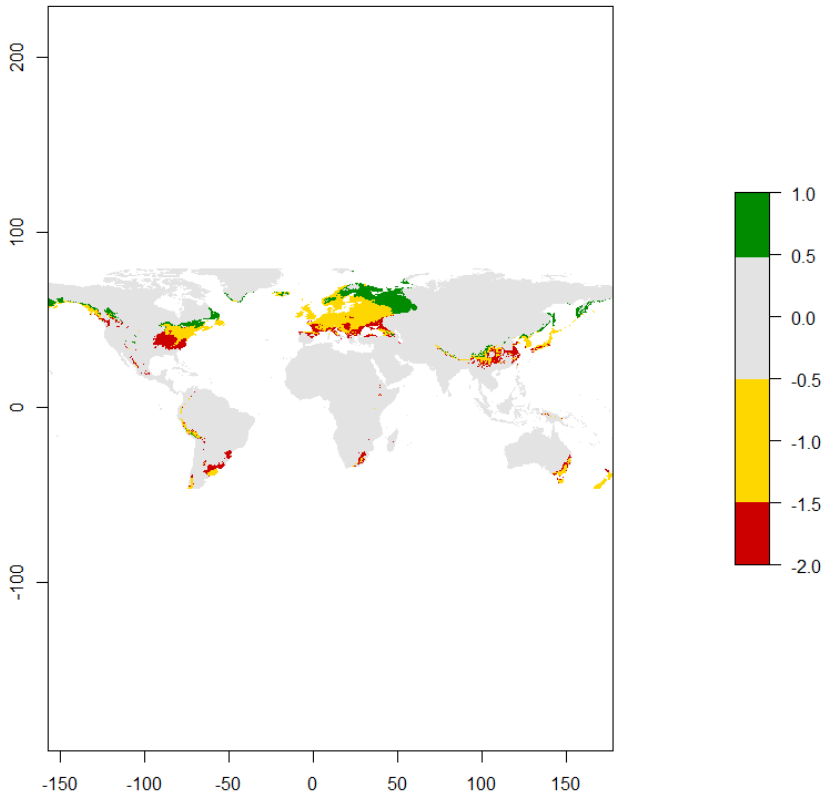


Figure 17: Change map of RCP 8.5 made in RStudio.

General interpretation of the future scenario's

In all the change maps the change between the present and future predictions is made visible for the different scenario's. In all the change maps the species will move to northern areas, these areas are coloured green. The habitat suitability is better than in the present data. With yellow the prediction is that the habitat suitability is lower than in the present. And red being that the habitat suitability is much lower than the present. In the southern hemisphere the species are predicted to decline in numbers. Probably because the species can't move to cooler places like in the northern hemisphere.

Discussion & conclusion

Buckwheat is currently present in relatively moist and cool temperate areas like Europe and North America. This is in line with the species biological characteristics and its habitat supported by literature. The hypothesis is that the species distribution changes to more northern areas and to higher altitudes, because this species is relatively sensitive to temperature change and precipitation. This is supported by the response curves (figure 6) and the table with the relative contributions of the environmental variables (figure 7). In the change maps the change in habitat suitability for this species from present to the different future predictions is made visible. It is shown that for all the different predictions the habitat suitability in almost all the areas is lower than the present. Also that the habitat suitability of northern areas is better than before. Species will move to northern areas, probably because of the temperature. The occurrences are in lower altitudes as in higher altitudes, but by looking at the change maps the species don't seem to move to higher altitudes as was expected before. A reason could be that the precipitation is different in higher altitudes, which is something that the buckwheat cannot deal with properly.

In conclusion, the predictions for the four different scenarios seem to be in line with the expectations and literature.

References

Hughes, H. D. and E. R. H. (1934). *Crop Production Principles and Practices.* New York.

Joshi, B. D. and R. S. P. (1991). Buckwheat in India. *National Bureau of Plant Genetic Resources*.

LI, & Q. (1992). Preliminary investigation on buckwheat origin in Yunnan, China. *Proc. 5th Intl. Symp. Buckwheat, Taiyuan*, 44–48.

Marshall, H. G. (1969). Isolation of Self-Fertile, Homomorphic Forms in Buckwheat, Fagopyrum sagittatum Gilib1. *Crop Science*, *9*(5), 651. https://doi.org/10.2135/cropsci1969.0011183x000900050044x

OHNISHI, O. (1988). Population genetics of cultivated common buckheat, Fagopyrum esculentum Moench. VII. Allozyme variability in Japan, Korea, and China. *The Japanese Journal of Genetics*, *63*(6), 507–522. https://doi.org/10.1266/jjg.63.507

OHNISHI, & O. (1995). Discovery of new Fagopyrum species and its implication for the studies of evolution of Fagopyrum and of the origin of cultivated buckwheat. *Proc. 6th Intl. Symp. Buckwheat, Ina*, 175–190.