The Future of Cultivated Rye (*Secale cereale*) for Different Climate Scenarios

Predicting present and future distribution of cultivated rye using MaxEnt species distribution modelling

MSc Course Methods in Biodiversity Analysis, Leiden University & Naturalis Biodiversity Center
Supervision: Dr. Rutger Vos & Dr. Maarten 't Zelfde
Name: Roel Wouters
Student ID: s1695894
Correspondence: WoutersRoel1@gmail.com
Date: December 2019

Table of Contents

Cover photo: Seeds from cultivated rye (*Secale cereal*e). iStockPhoto, derived from: Encyclopedia of Cultivated Plants: From Acacia to Zinnia, URL: https://ebrary.net/28094/environment/.

Introduction

Cultivated Rye

Rye (*Secale cereale,* Linnaeus, 1758) is a member of the grassy tribe Triticeae. Other cultivated grains like wheat (*Triticum)* and barley (*Hordeum*) are amongst its close relatives (Escobar et al., 2011)**.** Rye has three main utilisations, it is respectively used in i) bread, ii) whiskey and iii) as food for livestock. *Secale* is a small genus only consisting of three to four main species, however several subspecies are recognised (Al-Beyroutiová et al., 2016). Only *Secale cereale* and *Secale sylvestre* are considered domesticated (Tang et al., 2011). However, *Secale sylvestre* only has a small range in Eastern Europe and Western Asia and therefore, we use *Secale cereale* the remainder of this study. Rye grows in a broad range of climate conditions, it is well-adapted against colder and drier climates and is mostly found on sandy and rocky soils (Tang et al., 2011)**.** It is quite resilient compared to other cultivated grain species and is therefore widely distributed occurring on six continents, ranging from the Arctic circle to mountainous slopes in the tropics and subtropics (more information on its present distribution can be found under the method section). Due to its tolerance for colder climates and its resistance against diseases cultivated rye was one of the most important feeding crops during the Middle-Ages (Al-Beyroutiová et al., 2016). It occurs both as an annual as a perennial plant. Cultivated rye was believed to originate from Middle Asia. Maraci, Ozkan, & Bilgin (2018)**,** however, showed that for *Secale cereale* the genetic diversity in the Middle East, especially Syria, is the highest supporting the hypothesis that cultivated rye was domesticated and distributed from the Middle East. Because cultivated rye does, unlike other grains, generally well in colder and drier climates, an increase in temperature and a changing precipitation regime driven by climate change might have an important impact on its global distribution.

Climate Scenarios

To predict climate change in the next 50 years the International Panel for Climate Change (IPCC) adopted the so called Representative Concentration Pathway (RCP). This is measured as an increase in radiation. These four scenarios have been ultimately been established to predict global climate change over a 50 and 100 year time frame. Table 1 shows the four scenarios with the expected increase in temperature. For the remainder of this project, only the bioclimatic variables, in terms of temperature and precipitation, for the different RCPs have been used over a time period of 50 years.

Table 1) The different IPCC climate scenarios and the predicted temperature increase over 50 years. Note: the RCP scenarios are established for 100 years. We only focus on 50 years. Adapted from Pachauri et al. (2015).

Climate scenarios	Temperature increase (°C)
RCP2.6	1.0
RCP4.5	1.4
RCP6.0	1.3
RCP8.5	2.0

Species Distribution Modelling

Species Distribution Modelling (SDM) is an important tool in conservation biology, ecology and biodiversity research to predict future habitat suitability and distribution of species based on environmental data such as temperature, precipitation, soil type, etcetera. There are several SDM methods available. For this study we use the popular Maximum Entropy (MaxEnt) method.

Research Question

To get more insight in the future distribution of cultivated rye we propose our research question as "*What is the future habitat suitability and distribution of Secale cereale for the different IPCC climate scenarios using MaxExnt species distribution models?*"

Hypotheses

Here we hypothesise that with increasing temperature and changing precipitation regime cultivated rye will likely be distributed further north and south towards the Polar Regions because it does well in colder climates. Also, its abundance will decrease in mountainous habitats due to an increasing temperature. Furthermore, cultivated rye will likely be distributed more into drier areas because of its preference towards sandy soils and arid climate.

Methods

Present global occurrence data was derived from the open source GBIF database (DOI: https://doi.org/10.15468/dl.5octpg, obtained on December 3, 2019). The occurrence data is based on observations. The data was cleaned by selecting only observations with coordinates and selecting only living specimen (therefore leaving out museum specimen or processed plants). Also occurrences in the sea were removed as well as occurrence in the exact centre of a country. Furthermore, duplicates were taken out to prevent double counting of the same individual. This resulted in a total of 8505 observations. ArcGis was used to create a present occurrence map as shown in the result section. Environmental data was obtained from the open source WorldClim database. For both current and future conditions version 1.4 was used. For the current conditions only bioclimatic data was obtained with a resolution of 5 arc-minute (~10km grid). For the future conditions bioclimatic data was used for all four RCP scenarios, based on the IPCC 5th assessment report, with a similar resolution. Table 2 shows the bioclimatic variables used. Both the occurrence and environmental data were cleaned and prepared using Rstudio before running the SDM in the freely available MaxEnt software. The scripts used for data preparation were kindly provided by de (Vos & 't Zelfde, 2019).

Table 2) Description of the 19 bioclimatic variables as explained by the WorldClim database.

Code	Bioclimatic Variable
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter

BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

The scripts had to be slightly adjusted and personalised to the data of this study. First, bioclimatic variables were selected to include in the SDM. For this, an environmental autocorrelation test was performed. The resulting autocorrelation matrix can be found in Appendix 1. Furthermore, a multicolinearity test was performed, the outcomes can be found in Appendix 2. Based on these tests and based on the fact that, due to its natural habitat and climate, the biggest threat to cultivated rye might be temperature and precipitation increase, a total of five climate variables where selected for the SDM. Respectively, i) BIO2, ii) BIO5, iii) BIO8, iv) BIO13 and v) BIO18. For MaxEnt, the manual of Vos & ’t Zelfde (2019b) was used. We used the same settings with exception of the number of replicates, our analysis was done *in triplo*. The present occurrence data was extrapolated to create a SDM on a global scale. After running the SDM in MaxEnt, the output was validated using Area Under the Curve scores (AUC) and Jacknife. Finally, the MaxEnt output data was visualised in change maps using Rstudio according the manual (Vos & ’t Zelfde, 2019b). A complete overview of the workflow can be seen in Figure 1.

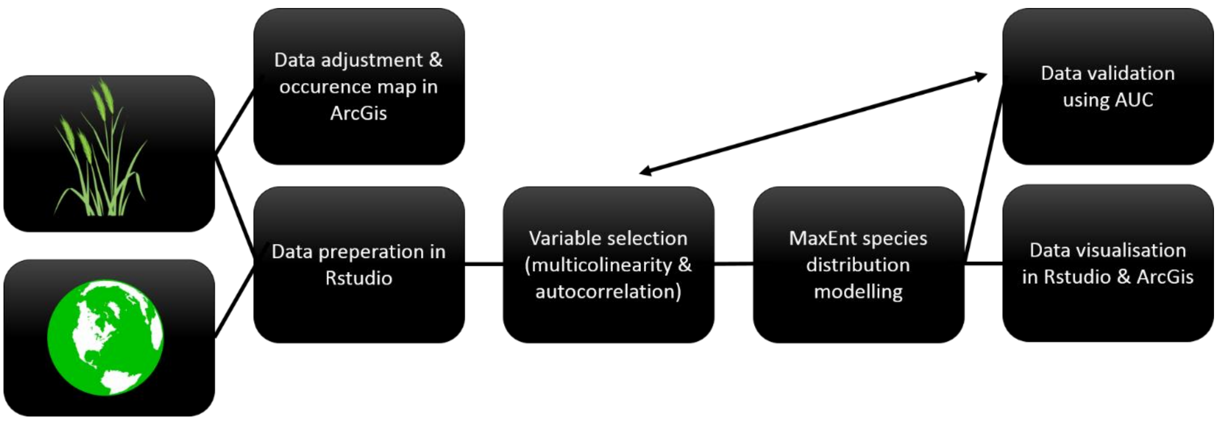

Figure 1) Workflow of the used methodology for Species Distribution Modelling. Occurrence data is symbolised as a grain icon, environmental data is symbolised as a globe icon.

Results

Present Distribution

First, we show an occurrence map with the current distribution of cultivated rye on a global scale (Figure 2a). It grows all over the world with a dense distribution in Europe and certain parts of the USA and Australia. It also occurs on mountain ranges such as the Himalaya and the Andes. Figure 2b depicts the present habitat suitability based on the occurrence data and environmental variables.

Present Global Distribution Map of Secale cereale

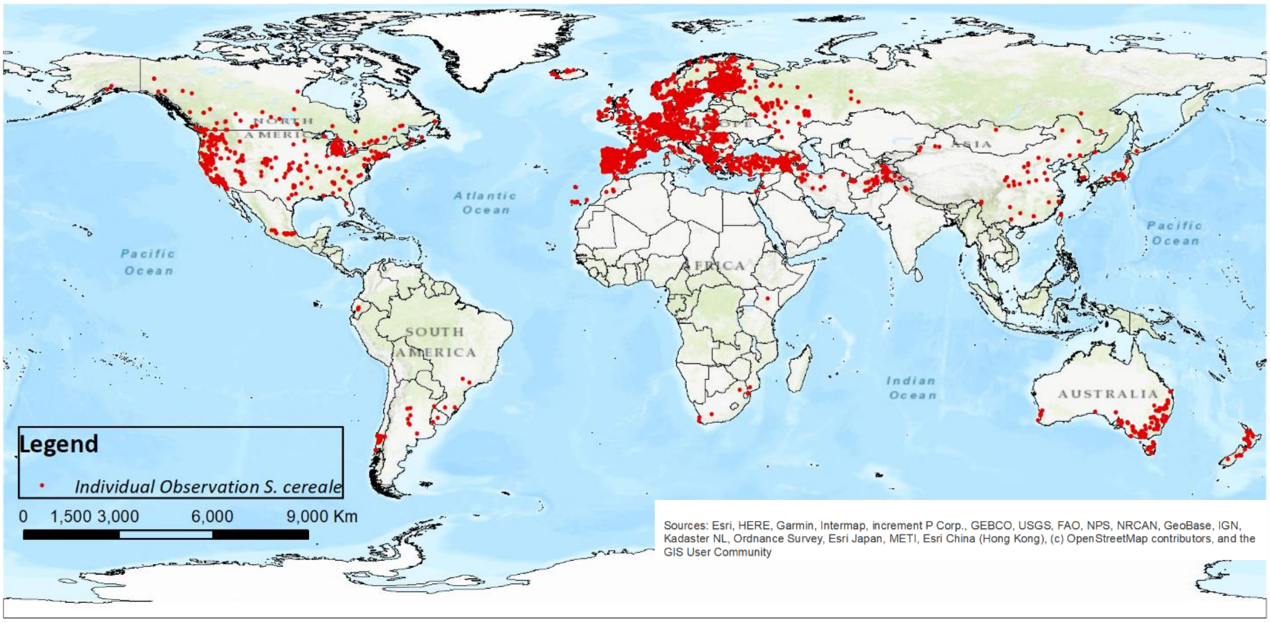

Figure 2a) Global distribution of Secale cereal based on occurrence data from open source database gbif.

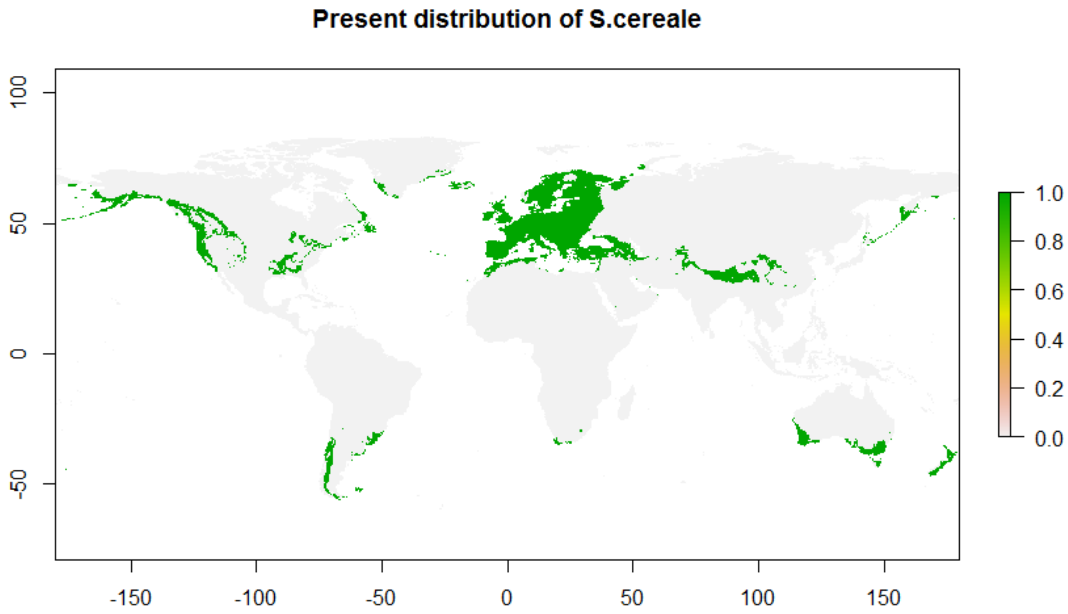

Figure 2b) The present habitat suitability of Secale cereal based on present occurrence and environmental data.

Model Diagnostics

After running the SDM in MaxEnt model diagnostics were examined to determine the reliability of the model. One of the diagnostics used for reliability is the AUC score. If AUC > 0.8 we consider the model reliable. Our model performed with an AUC score of 0.845. Note that this is the mean AUC of our three models. Apendix 3 shows the matching graph. Furthermore the variable importance was examined using Jacknife. Table 3 shows BIO4 (temperature seasonality), BIO 8 (mean temperature of wettest quarter) and BIO13 (precipitation of the wettest month) make up more than 80 percent, indicating these variables contributed the most to the SDM. This is further illustrated by Figure 3 showing the response curves of the variables, indicating how each bioclimatic variable affects the SDM.

Table 3) Variable importance for each of the used bioclimatic variable in the SDM. Derived from MaxEnt output.

Variable	Percent contribution	Permutation importance
_Bio04	47.6	60.5
_Bio08	30.5	22.7
_Bio13	15.8	10.4
_Bio02	5.1	4.1
_Bio18	1.1	2.4

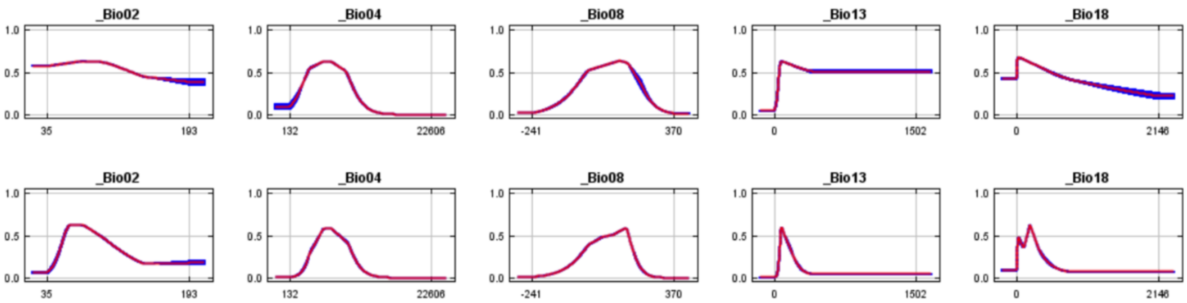

Figure 3) Response curves of the showing how each of the environmental variable used influences the MaxEnt SDMl. Derived from MaxEnt output.

RCP2.6

For RCP2.6 there is only a slight difference noticeable when comparing the present and future distribution in terms of habitat suitability of cultivated rye (Figure 4a). The loss and gain of suitable habitat is depicted in Figure 4b. We can observe that there is a slight increased suitability moving up north in USA and Europe as well as in the Himalaya mountain range in Central Asia. Habitat loss is most prominent in the Eastern part of Europe as well as Central USA and Southern Australia.

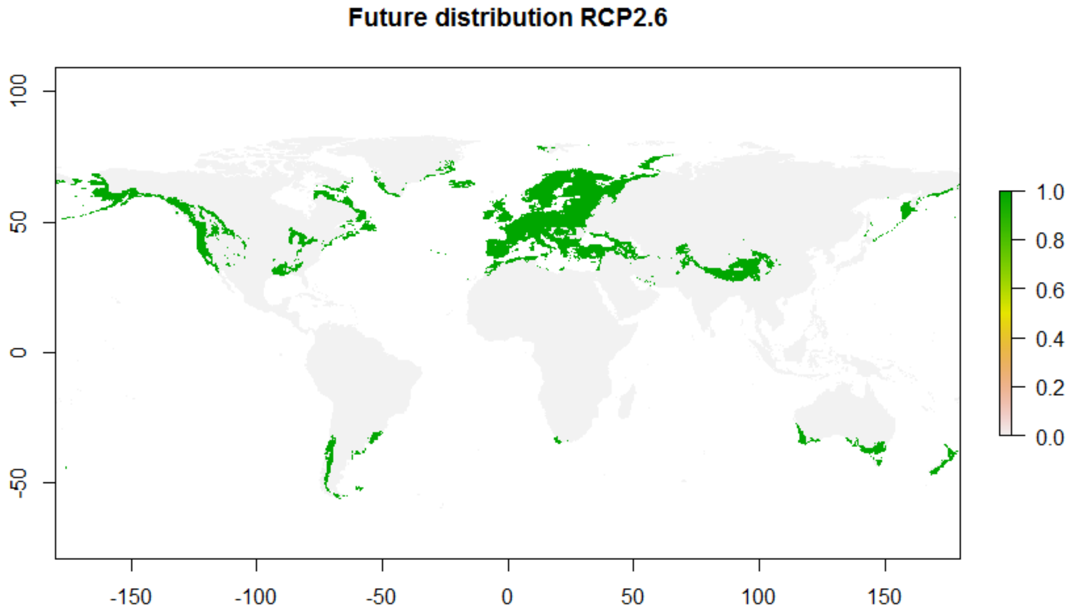

Figure 4a) Future distribution, based on habitat suitability of Secale cereale in scenario RCP2.6

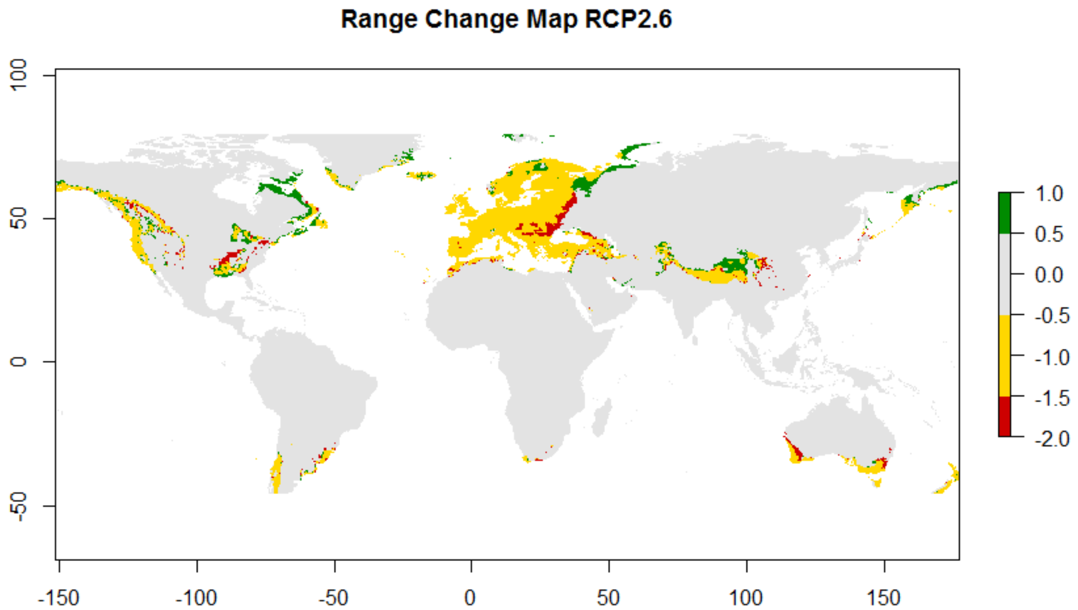

Figure 4b) Range change map, showing the habitat suitability loss (red) and gain (green) for Secale cereale. Areas that remain suitable are marked yellow.

RCP4.5

For RCP4.5 similar results are obtained as for RCP2.6, indicating these two scenarios don't differ much from each other (Figure 5a). We can observe slightly more habitat suitability loss in the USA and Australia. However, the loss of habitat suitability in Europe remains much smaller compared to the previous RCP2.6 (Figure 5b).

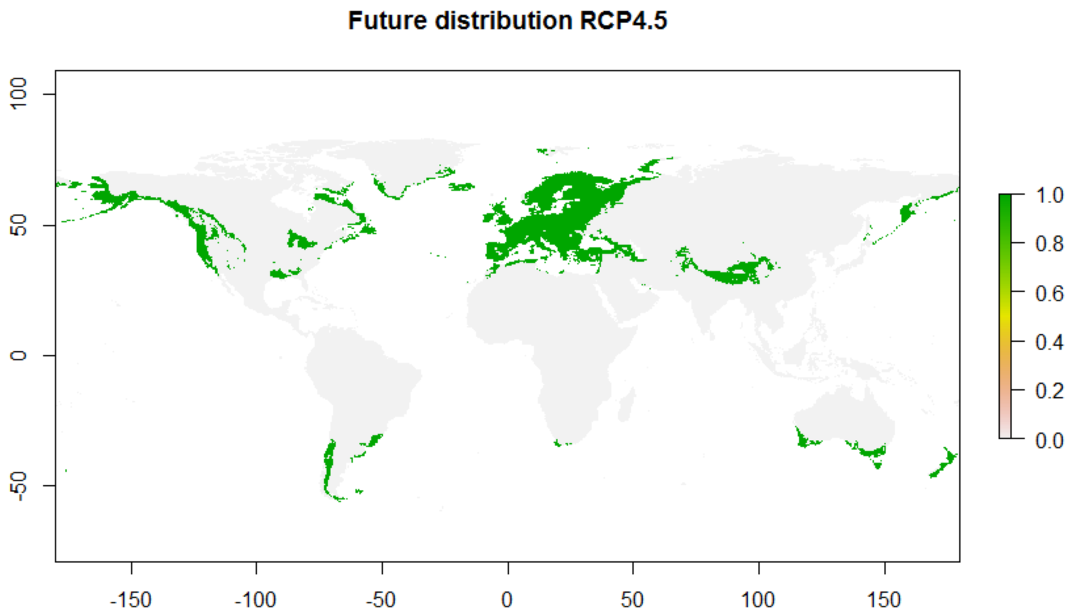

Figure 5a) Future distribution, based on habitat suitability of Secale cereale in scenario RCP4.5

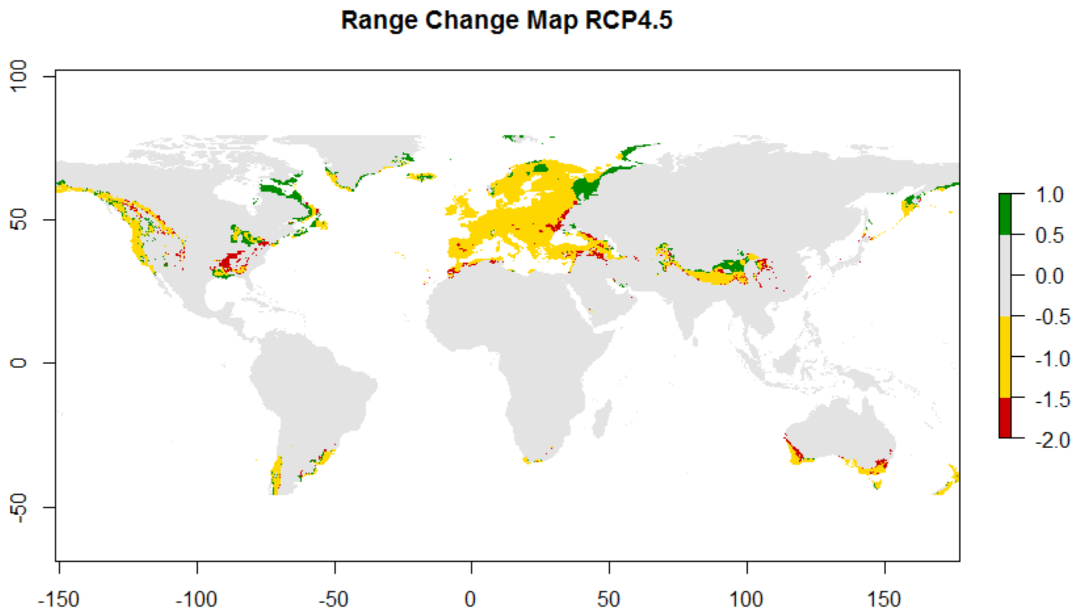

Figure 5b) Range change map, showing the habitat suitability loss (red) and gain (green) for Secale cereale. Areas that remain suitable are marked yellow.

RCP6.0

Figure 6 shows a similar trend to the previous RCP scenarios. However more habitat suitability loss is observed in Central and Eastern Europe and on the other hand more habitat suitable area is gained in Northern USA and Greenland.

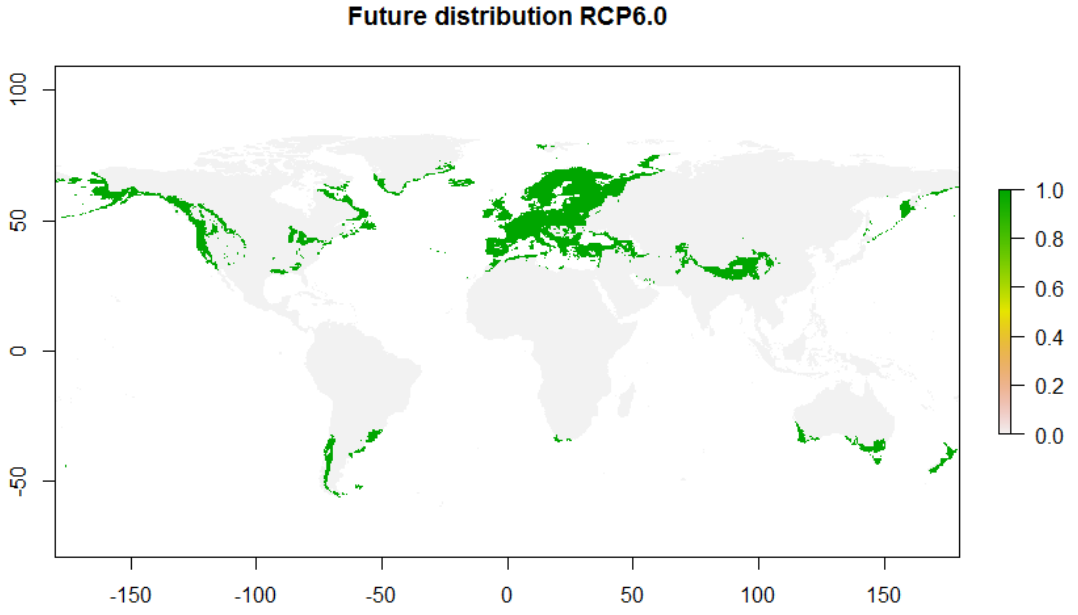

Figure 6a) Future distribution, based on habitat suitability of Secale cereale in scenario RCP6.0

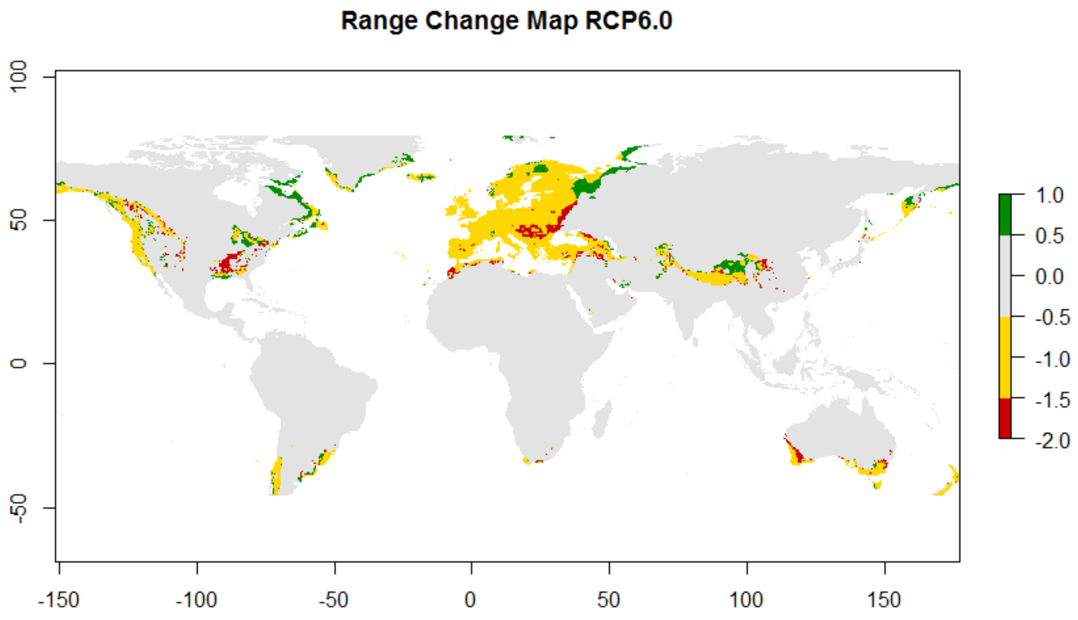

Figure 6b) Range change map, showing the habitat suitability loss (red) and gain (green) for Secale cereale. Areas that remain suitable are marked yellow.

RCP8.5

In contradiction to the previous RCP scenarios, RCP8.5 predicts more habitat gain and remarkably less habitat loss in terms of suitability. Figure 7 shows an increase in habitat suitability in Northern USA and Canada.

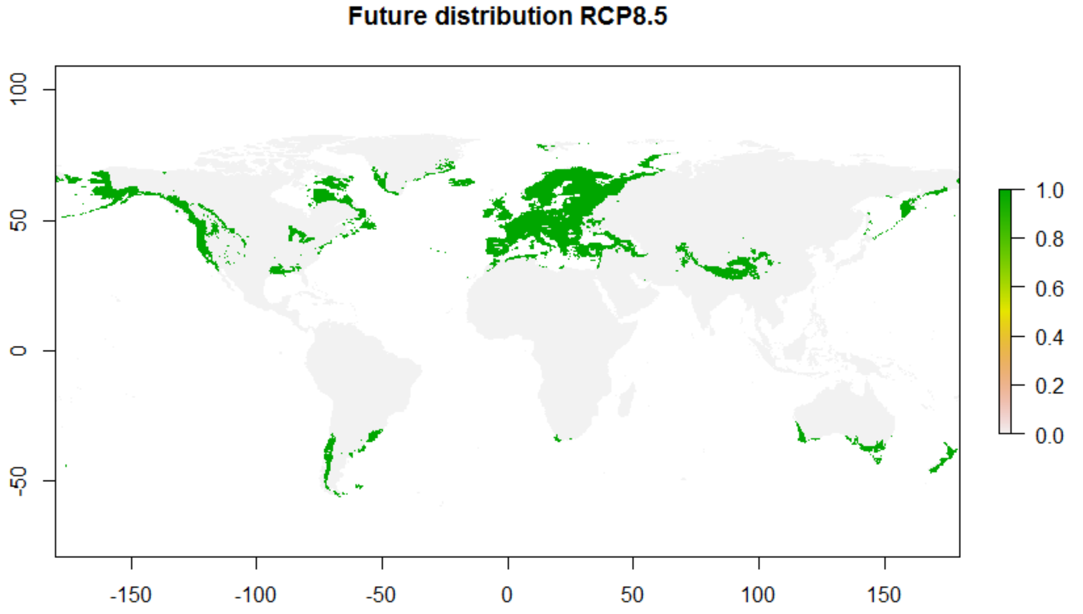

Figure 7a) Future distribution, based on habitat suitability of Secale cereale in scenario RCP8.0

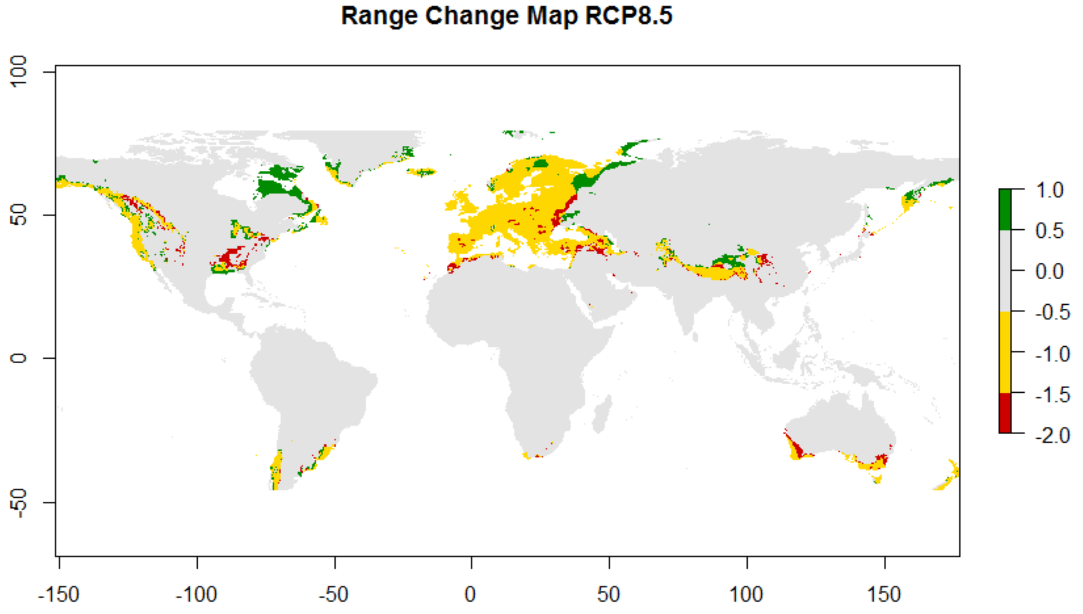

Figure 7b) Range change map, showing the habitat suitability loss (red) and gain (green) for Secale cereale. Areas that remain suitable are marked yellow.

Discussion & Conclusion

Our results show the changing distribution (habitat suitability) of cultivated rye is driven by (low) temperature seasonality and temperature and precipitation in the wettest month. Furthermore the models show a gain in habitat suitability in the Northern parts of the USA, Canada and Europe, which is more or less consistent over the different RCP scenarios. Habitat loss is most severe in central USA and Europe. There is not much change in the mountainous habitats. It has to be noted that for this SDM differences in altitude were not taken into account and therefore, the modelled effects of climate change on cultivated rye habitat suitability in mountainous areas might not reflect reality. However, overall, the global distribution of habitat suitability doesn't deviate that much from the present situation indicating that the used environmental variables are probably not the best indicators for predicting of habitat suitability for cultivated rye. For example, changing soil types will most likely also play an important role, especially because cultivated rye naturally occurs on sandy and rocky soils. Nevertheless cultivated rye is a generalist species doing remarkably well in many temperate and colder climates and habitats all around the globe, this might also partially explain the resilient character of rye the SDM depicts for all scenarios. It must also be taken into account that especially for Africa, South America and Australia there are not many observations, resulting in a possible overestimation of these regions in our SDM, making these regions less reliable. Our model performed well indicated by the AUC score and the response curves. To obtain a more quantitative understanding of the habitat suitability change the output maps could be analysed in ArcGIS. This would give the possibilities to measure the area of specific regions where habitat suitability loss/gain play an important role. Cultivated rye is a domesticated crop that is produced and maintained by human. It is an important food source for human and cattle. Therefore the global effects of climate change might not be so apparent after all since human will come up with methods to remain cultivating rye (e.g. placing them in climate controlled greenhouses and artificial irrigation).

In conclusion our SDM shows temperature seasonality and temperature and precipitation in the wettest month are the most influential environmental variables in the depicted change in habitat suitability for cultivated rye. Predictions are consistent between different scenarios. The SDM shows a loss in habitat suitability in Central USA and Central Europe as well as Southern Australia. It shows a habitat suitability increase further up north in USA, Canada and Europe. In general, the predicted habitat suitability change does not deviate from the present distribution. Therefore other variables such as soil types, altitudes and diseases are believed to have an important effect on SDM and should be taken into account in future research. Despite the variable selection, our model performed well and due to its resilient nature, tolerating a wide range of climates and habitats the effects of climate change might not greatly affect the global distribution of cultivated rye in the future.

Acknowledgements

I would like to acknowledge Rutger Vos and Maarten 't Zelfde for their supervision, the supply of course material and lab space (GIS lab). I would like to thank Nuno De Mesquita César de Sá and Rosaleen March from the Institute of Environmental Sciences, Leiden University for their help in preparations and theoretical knowledge on species distribution modelling. I would like to thank Jordy van der Beek, Sabine Werndlij, Pip van Aalst, Nathalie Winkster and Katrijn De Bock for their (scientific) help and support during this project. Special thanks go to Koen Rurrenga who has been of great help in solving multiple script errors and general insights in script programming.

References

Al-Beyroutiová, M., Sabo, M., Sleziak, P., Dusinsky, R., Bircák, E., Hauptvogel, P., … Svec, M. (2016). Evolutionary relationships in the genus Secale revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution*. https://doi.org/10.1007/s00606-016-1318-2

Escobar, J. S., Scornavacca, C., Cenci, A., Guilhaumon, C., Santoni, S., Douzery, E. J. P., … David, J. (2011). Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evolutionary Biology*, *181*(June). https://doi.org/doi:10.1186/1471-2148-11-181

Maraci, O., Ozkan, H., & Bilgin, R. (2018). Phylogeny and genetic structure in the genus Secale. *PLoS ONE*, 1–21. https://doi.org/https://doi.org/10.1371/journal.pone.0200825

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., … van Ypersele, J.-P. (2015). *Climate Change 2014 Synthesis Report*. Retrieved from https://www.ipcc.ch/report/ar5/syr/

Tang, Z. X., Ross, K., Ren, Z. L., Yang, Z. J., Zhang, H. Y., Chikmawati, T., … Gusafson, J. P. (2011). Secale. In M. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources: Cereals* (pp. 367–395). Springer. https://doi.org/DOI 10.1007/978-3-642-14228-4_8

Vos, R., & ’t Zelfde, M. (2019a). Model your chosen species ’ habitat suitability under present and future climate conditions. Retrieved from https://github.com/naturalis/mebioda/blob/master/doc/week2/w2d5/Mebioda_PracticalManual_2019.pdf

Vos, R., & ’t Zelfde, M. (2019b). R code. Retrieved from https://github.com/naturalis/mebioda/tree/master/doc/week2/w2d5

Appendices

Appendix 1

Environmental autocorrelation matrix, all values <-0,7 and >0,7 are considered highly correlated and were therefore not regarded in the SDM (not that all values are separated by commas).

	Bio01	Bio02	Bio03	Bio04	Bio05	Bio06	Bio07	Bio08	Bio09	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19
Bio01	1	0,4817813	0,8356903	-0,841042	0,8901112	0,9684318	-0,748875	0,8007466	0,9380601	0,9322349	0,9802878	0,3724381	0,4477961	0,0537338	0,3561510	0,4366635	0,0810442	0,2112005	0,245991799756634
Bio02	0,4817813	1	0,3566494	-0,188138	0,6789156	0,3170626	0,0296238	0,4934904	0,4049249	0,5851736	0,3893764	-0,281186	-0,139632	-0,394765	0,5057387	-0,160252	-0,393916	-0,235395	-0,26614324919305
Bio03	0,8356903	0,3566494	1	-0,891803	0,6023346	0,8842433	-0,832983	0,6308005	0,8032294	0,6515833	0,8875129	0,5631015	0,5763217	0,2244955	0,2732310	0,5764946	0,2583143	0,3475415	0,430141279131851
Bio04	-0,841042	-0,188138	-0,891803	1	-0,515367	-0,939213	0,9729973	-0,503059	-0,864151	-0,590148	-0,930264	-0,552025	-0,565280	-0,242372	-0,188340	-0,566301	-0,272653	-0,332396	-0,402046263329283
Bio05	0,8901112	0,6789156	0,6023346	-0,515367	1	0,7588517	-0,372269	0,8311324	0,7878028	0,9872537	0,7894204	0,0995771	0,2133229	-0,150198	0,4178166	0,1947111	-0,132406	-0,000422	0,0435790404312983
Bio06	0,9684318	0,3170626	0,8842433	-0,939213	0,7588517	1	-0,886950	0,6951163	0,9480876	0,8226258	0,9959415	0,4830896	0,5216598	0,1668218	0,2591522	0,5163036	0,1970341	0,2735617	0,354978511245647
Bio07	-0,748875	0,0296238	-0,832983	0,9729973	-0,372269	-0,886950	1	-0,401196	-0,792438	-0,472193	-0,859488	-0,617839	-0,592140	-0,344258	-0,073013	-0,597706	-0,374696	-0,390156	-0,474979523416099
Bio08	0,8007466	0,4934904	0,6308005	-0,503059	0,8311324	0,6951163	-0,401196	1	0,5927532	0,8538311	0,7214019	0,2433307	0,3607816	-0,068973	0,4365083	0,3423507	-0,047121	0,2448705	0,0664828594914215
Bio09	0,9380601	0,4049249	0,8032294	-0,864151	0,7878028	0,9480876	-0,792438	0,5927532	1	0,8293472	0,9492273	0,3694236	0,4059601	0,1075223	0,2497129	0,4003775	0,1343783	0,1408274	0,305383653896425
Bio10	0,9322349	0,5851736	0,6515833	-0,590148	0,9872537	0,8226258	-0,472193	0,8538311	0,8293472	1	0,8449214	0,1869433	0,2894369	-0,080269	0,3990745	0,2719396	-0,059734	0,0813680	0,105355856065772
Bio11	0,9802878	0,3893764	0,8875129	-0,930264	0,7894204	0,9959415	-0,859488	0,7214019	0,9492273	0,8449214	1	0,4495136	0,5052793	0,1224761	0,3074580	0,4980072	0,1518441	0,2557353	0,313525761861801
Bio12	0,3724381	-0,281186	0,5631015	-0,552025	0,0995771	0,4830896	-0,617839	0,2433307	0,3694236	0,1869433	0,4495136	1	0,8963155	0,7060779	-0,183289	0,9220947	0,7414334	0,7926200	0,751350665875258
Bio13	0,4477961	-0,139632	0,5763217	-0,565280	0,2133229	0,5216598	-0,592140	0,3607816	0,4059601	0,2894369	0,5052793	0,8963155	1	0,3900511	0,1290201	0,9925785	0,4272316	0,7380922	0,581979086234078
Bio14	0,0537338	-0,394765	0,2244955	-0,242372	-0,150198	0,1668218	-0,344258	-0,068973	0,1075223	-0,080269	0,1224761	0,7060779	0,3900511	1	-0,525772	0,4262831	0,9937281	0,5529509	0,669234247726409
Bio15	0,3561510	0,5057387	0,2732310	-0,188340	0,4178166	0,2591522	-0,073013	0,4365083	0,2497129	0,3990745	0,3074580	-0,183289	0,1290201	-0,525772	1	0,0844942	-0,524429	-0,114997	-0,279377781052584
Bio16	0,4366635	-0,160252	0,5764946	-0,566301	0,1947111	0,5163036	-0,597706	0,3423507	0,4003775	0,2719396	0,4980072	0,9220947	0,9925785	0,4262831	0,0844942	1	0,4636185	0,7568002	0,607265589588075
Bio17	0,0810442	-0,393916	0,2583143	-0,272653	-0,132406	0,1970341	-0,374696	-0,047121	0,1343783	-0,059734	0,1518441	0,7414334	0,4272316	0,9937281	-0,524429	0,4636185	1	0,5761129	0,696798875928033
Bio18	0,2112005	-0,235395	0,3475415	-0,332396	-0,000422	0,2735617	-0,390156	0,2448705	0,1408274	0,0813680	0,2557353	0,7926200	0,7380922	0,5529509	-0,114997	0,7568002	0,5761129	1	0,366659067628543
Bio19	0,2459917	-0,266143	0,4301412	-0,402046	0,0435790	0,3549785	-0,474979	0,0664828	0,3053836	0,1053558	0,3135257	0,7513506	0,5819790	0,6692342	-0,279377	0,6072655	0,6967988	0,3666590	1

Appendix 2

VIF multicolinearity test output in Rstudio. A VIF score <10 is considered generally reliable, a VIF score <5 is considered highly reliable.

> vif(df.stk.AOI, maxobservations=nrow(df.stk.AOI))

Variables	VIF
Bio02	2.206812
Bio05	5.213340
Bio08	4.020450
Bio13	2.603123
Bio18	2.564374

Appendix 3

Area Under the Curve (AUC) graph showing an AUC score of 0.845 on our SDM.